[redacted]

AUSTRALIAN JOURNAL OF PHYSICS

VOLUME 13, NUMBER 1, MARCH 1960

REGISTERED IN AUSTRALIA FOR TRANSMISSION BY POST AS A PERIODICAL

AUSTRALIAN JOURNAL OF PHYSICS

Published by the Commonwealth Scientific and Industrial Research Organization. Volumes 1 to 5 of the Australian Journal of Physics and the Australian Journal of Chemistry issued as the Australian Journal of Scientific Research, Series A: Physical Sciences. Issued quarterly, £2 per annum.

BOARD OF STANDARDS

The Board of Standards for this Journal is appointed by the Commonwealth Scientific and Industrial Research Organization and the Australian Academy of Science and consists of Professor Sir Macfarlane Burnet (Chairman), Dr. N. S. Noble (Editor), Professor N. S. Bayliss, Dr. J. L. Pawsey, Professor W. P. Rogers, and Professor J. S. Turner.

ADVISORY COMMITTEE

Acceptance of papers for this Journal is in the hands of an Advisory Committee appointed by the Board of Standards in consultation with the Institute of Physics (Australian Branch) and consisting of Dr. N. S. Noble (Chairman and Editor), Dr. G. H. Briggs, Professor L. G. H. Huxley, Dr. J. L. Pawsey, and Professor H. C. Webster.

OTHER JOURNALS PUBLISHED BY C.S.I.R.O.

Australian Journal of Chemistry Australian Journal of Biological Sciences Australian Journal of Applied Science	Issued quarterly, £2 per annum
Australian Journal of Agricultural Research	Issued six times a year, £2 per annum
Australian Journal of Marine and Freshwater Research Australian Journal of Botany Australian Journal of Zoology	Not issued at regular intervals, 10/- per issue

The Organization is a signatory to the Fair Copying Declaration, details of which may be obtained from the office of the Royal Society, London.

All enquiries and manuscripts should be forwarded to:

The Editor,
Australian Journal of Physics,
Commonwealth Scientific and Industrial Research Organization,
314 Albert Street, East Melbourne, C.2, Victoria

MELBOURNE

1127 L

GENERAL METHOD OF EXACT SOLUTION OF THE CONCENTRATION-DEPENDENT DIFFUSION EQUATION

By J. R. Philip*

[*Manuscript received November* 11, 1959]

Summary

Only three forms of $D(\theta)$ have previously been known to yield exact solutions of the equation

$$\frac{\partial\theta}{\partial t}=\frac{\partial}{\partial x}\left(D(\theta)\frac{\partial\theta}{\partial x}\right),$$

subject to the conditions $\theta=0,\ x>0,\ t=0$; $\theta=1,\ x=0,\ t>0$. The present paper reports a general method of establishing a very large class of $D(\theta)$ functions which yield exact solutions. A similar method enables exact solutions of the same equation subject to the conditions $\theta=0,\ x>0$, and $\theta=1,\ x<0,\ t=0$; $\int_0^1 x\mathrm{d}\theta=0,\ t\geqslant 0$. In this case also a very large class of $D(\theta)$ functions yield exact solutions. Examples are given for both cases.

Many of the exact solutions which are most readily found tend to lead to zero or infinite values of D at one or two points of the θ-range. Means of avoiding this difficulty are devised. Practical use of the method is discussed.

I. Introduction

Hitherto, the mathematics of concentration-dependent diffusion has depended almost exclusively on numerical analysis. According to Crank (1956, p. 166), the only known " formal solutions " in concentration-dependent diffusion are due to Fujita (1952*a*, 1952*b*, 1954). They provide solutions of the equation

$$\frac{\partial\theta}{\partial t}=\frac{\partial}{\partial x}\left(D(\theta)\frac{\partial\theta}{\partial x}\right), \qquad \text{(1.1)}$$

subject to the conditions

$$\left.\begin{array}{l}\theta=0,\ x>0,\ t=0\ ;\\ \theta=1,\ x=0,\ t>0\ ;\end{array}\right\} \qquad \text{(1.2)}$$

and are for the following $D(\theta)$ functions :

$$D=D_0/(1-\lambda\theta)\ ;\quad D=D_0/(1-\lambda\theta)^2\ ;\quad D=D_0/(1+2a\theta+b\theta^2).$$

Philip (1955) showed how (**1.1**) subject to (**1.2**) could be solved analytically for D, an n-step function of θ. However, the method is cumbersome when n is large.

This paper presents a very general method of obtaining exact solutions of (**1.1**) subject to (**1.2**). It embraces all possible exact solutions, including as particular cases those of Fujita and those implicit in Philip (1955). The method is also applicable to all possible exact solutions of (**1.1**) subject to the following conditions, which are also of practical importance:

---

* Division of Plant Industry, C.S.I.R.O., Canberra.

$$\theta=0,\ x>0 \text{ and } \theta=1,\ x<0,\ t=0;\quad \int_0^1 x\,\mathrm{d}\theta=0,\ t\geqslant 0. \qquad \ldots . \ (\mathbf{1.3})$$

Extension of the results of this paper to the slightly more general cases of (**1.2**) and (**1.3**) with $\theta=\theta_n$, $\theta=\theta_0$, in place of $\theta=0$, $\theta=1$, is trivial, involving only the introduction of the linear transformation $\theta_*=(\theta-\theta_n)/(\theta_0-\theta_n)$ and solution in terms of θ_*.

II. Exact Solutions and Exact Functions

The use of the adjective "exact" to distinguish solutions obtainable in terms of the functions of analysis from those obtainable only numerically is well established (e.g. Carslaw and Jaeger 1959, p. 91). This usage is convenient and unambiguous, though no definition of the term seems to have been given in the literature.

For the purposes of the present work, it is desirable to define an *exact function* of a real variable. We say that y is an exact function of the real variable x in some interval of x, provided y can be specified without approximation for all values of x in the given interval by means of a finite number of explicit formulae.

Further, we may then define an *exact solution* as a solution expressible in the form of an exact function.

III. General Principles of the Method

It is well known (cf. Philip 1955) that the substitution

$$\varphi=xt^{-\frac{1}{2}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (\mathbf{3.1})$$

enables (**1.1**) subject to either (**1.2**) or (**1.3**) to be reduced to

$$\int_0^\theta \varphi\,\mathrm{d}\theta=-2D\,\mathrm{d}\theta/\mathrm{d}\varphi, \qquad \ldots\ldots\ldots\ldots\ldots \ (\mathbf{3.2})$$

that is,

$$D=-\tfrac{1}{2}\mathrm{d}\varphi/\mathrm{d}\theta.\int_0^\theta \varphi\,\mathrm{d}\theta. \qquad \ldots\ldots\ldots\ldots \ (\mathbf{3.3})$$

It follows that the solution of (**1.1**), subject to either (**1.2**) or (**1.3**), $\varphi(\theta)$, exists in exact form, so long as $D(\theta)$ is of the form

$$D=-\tfrac{1}{2}\mathrm{d}F/\mathrm{d}\theta.\int_0^\theta F\,\mathrm{d}\theta, \qquad \ldots\ldots\ldots\ldots \ (\mathbf{3.4})$$

where F is any single-valued exact function of θ in the interval $0\leqslant\theta\leqslant 1$, which satisfies certain, not very restrictive, conditions, which we indicate in the following paragraphs.†

---

† Note that D is expressible in terms of known functions if F (as well as satisfying these conditions) is integrable in terms of known functions.

(i) *Conditions imposed by Conditions governing* (**1.1**).—Evidently, the exact solution for $D(\theta)$ of form (**3.4**) is simply $\varphi=F$. F must therefore satisfy the conditions to be imposed on (φ in) equation (**3.2**). It follows that, for (**1.1**) subject to (**1.2**),

$$F(1)=0, \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{3.5})$$

and that, for (**1.1**) subject to (**1.3**),

$$\int_0^1 F\mathrm{d}\theta=0. \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{3.6})$$

(ii) *Conditions imposed by Requirement that* D *Exists.*—For D to exist for all values of θ in the range $0\leqslant\theta\leqslant1$, it is necessary that $\int_0^\theta F\mathrm{d}\theta$ and $\mathrm{d}F/\mathrm{d}\theta$ exist throughout this θ-range. (However, if a finite number of discontinuities in D are allowable, or if D is permitted to be infinite at a finite number of points in the range $0<\theta\leqslant1$, $\mathrm{d}F/\mathrm{d}\theta$ may either not exist or be infinite at the appropriate finite number of points in the θ-range.)

(iii) *Condition imposed by Requirement that* $\mathrm{D}\geqslant0$.—The flux of the diffusing entity in the sense x positive, Q, is equal to $-D\partial\theta/\partial x=-t^{-\frac{1}{2}}D\mathrm{d}\theta/\mathrm{d}\varphi$. Now both (**1.2**) and (**1.3**) describe phenomena in which for $D\geqslant0$, $Q\geqslant0$. It follows that $\mathrm{d}\theta/\mathrm{d}\varphi\leqslant0$ in both cases. Accordingly we have, for the two cases we consider, the further condition

$$\mathrm{d}F/\mathrm{d}\theta\leqslant0 \quad (0\leqslant\theta\leqslant1). \qquad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{3.7})$$

IV. SOLUTIONS OF (1.1), (1.2)

It is seen that exact solutions of (**1.1**) subject to (**1.2**) may be established at will, simply by selecting F functions satisfying the conditions set out in Section III. Some typical results of this elementary process are given in Table 1.

It will be observed that the D functions in this table tend to have $\lim_{\theta\to0} D(\theta)$ either zero or infinite.* It would be a serious limitation on the generality of the present methods if none of our exact solutions were for $D(0)$ non-zero, but finite.† Accordingly, we now give some attention to this question.

A well-known result in linear diffusion (Carslaw and Jaeger 1959, p. 59) for phenomena governed by (**1.2**) is equivalent, from the present viewpoint, to the statement that

$$\text{for } F=2D_0^{\frac{1}{2}}\,\text{inverfc}\,\theta,\ D=\text{constant}=D_0. \qquad \ldots\ldots \quad (\mathbf{4.1})$$

The notation inverfc denotes the inverse of erfc (Philip 1955).

---

[*] From this point on we shall write $D(0)$ for $\lim_{\theta\to0} D(\theta)$. Note that we use the form $\lim_{\theta\to0} D(\theta)$ since, strictly, the value of D at $\theta=0$ is irrelevant to the phenomena. We are concerned only with values of $D(\theta)$ in the interval $0<\theta\leqslant1$.

[†] This is not to imply that no practical interest attaches to cases with $D(0)=0$. For a discussion of such cases and certain applications see Philip (1956, 1957).

We therefore put

$$F=2D_0^{\frac{1}{2}}\ \text{inverfc}\ \theta+f(\theta)\ ;\quad D_0\ \text{finite, non-zero} \qquad \ldots\ldots (4.2)$$

in (**3.4**), and obtain

$$D=D_0+\tfrac{1}{2}D_0^{\frac{1}{2}}\left(\frac{2}{B}\int_0^\theta f\mathrm{d}\theta-B\frac{\mathrm{d}f}{\mathrm{d}\theta}\right)-\tfrac{1}{2}\frac{\mathrm{d}f}{\mathrm{d}\theta}\cdot\int_0^\theta f\mathrm{d}\theta, \qquad \ldots\ldots (4.3)$$

where

$$B(\theta)=2\pi^{-\frac{1}{2}}\exp\left[-(\text{inverfc}\ \theta)^2\right]. \qquad \ldots\ldots (4.4)$$

TABLE 1

SOME SIMPLE CASES OF EXACT SOLUTIONS OF (**1.1**), (**1.2**)

No.	D	φ	Remarks
1	$\frac{1}{2}n\theta^n[1-\theta^n/(1+n)]$	$1-\theta^n$	$n>0$
2	$\dfrac{n}{2(n+1)}[(1-\theta)^{n-1}-(1-\theta)^{2n}]$	$(1-\theta)^n$	$n>0$
3	$\frac{1}{2}n\theta^{-n}\left[\dfrac{\theta^{-n}}{1-n}-1\right]$	$\theta^{-n}-1$	$0<n<1$
4	$\frac{1}{2}\sin^2\frac{1}{2}\pi\theta$	$\cos\frac{1}{2}\pi\theta$	
5	$\frac{1}{2}\cos\frac{1}{2}\pi\theta[\cos\frac{1}{2}\pi\theta+\frac{1}{2}\pi\theta-1]$	$1-\sin\frac{1}{2}\pi\theta$	
6	$\frac{1}{2}\left(\dfrac{\theta\cos^{-1}\theta+1}{\sqrt{(1-\theta^2)}}-1\right)$	$\cos^{-1}\theta$	
7	$\frac{1}{2}\left(\dfrac{\frac{1}{2}\pi-(1-\theta)\sin^{-1}(1-\theta)}{\sqrt{(2\theta-\theta^2)}}-1\right)$	$\sin^{-1}(1-\theta)$	
8	$\frac{1}{4}\theta\left(\theta+\dfrac{\sin^{-1}\theta}{\sqrt{(1-\theta^2)}}\right)$	$\sqrt{(1-\theta^2)}$	
9	$\frac{1}{8}\sin\pi\theta(\pi\theta+\sin\pi\theta)$	$\cos^2\frac{1}{2}\pi\theta$	
10	$\frac{1}{16}\sin^2\pi\theta(5+\cos\pi\theta)$	$\cos^3\frac{1}{2}\pi\theta$	
11	$\frac{1}{2}\mathrm{e}^{n\theta}[1-\mathrm{e}^{n\theta}+n\theta\mathrm{e}^n]$	$\mathrm{e}^n-\mathrm{e}^{n\theta}$	$n>0$
12	$\frac{1}{2}\mathrm{e}^{-n\theta}[1-\mathrm{e}^{-n\theta}-n\theta\mathrm{e}^{-n}]$	$\mathrm{e}^{-n\theta}-\mathrm{e}^{-n}$	$n>0$
13	$\frac{1}{2}(1-\log\theta)$	$-\log\theta$	
14	$\dfrac{n}{2\theta}(-\log\theta)^{n-1}$ $\times\{\Gamma(n+1)-\Gamma(-\log\theta,\ n+1)\}$	$(-\log\theta)^n$	$n>0$. $\Gamma(x,p)$ denotes the incomplete Γ-function $\int_0^x \mathrm{e}^{-x}x^{p-1}\mathrm{d}x$.

This result depends on the following relationships, which are readily established :

$$\frac{\mathrm{d}}{\mathrm{d}\theta}(\text{inverfc}\ \theta)=-\frac{1}{B}\ ;\quad \int_0^\theta \text{inverfc}\ \theta\ \mathrm{d}\theta=\tfrac{1}{2}B. \qquad \ldots (4.5)$$

The properties of inverfc θ and $B(\theta)$ are treated in more detail in Philip (1960). Tables of these two functions are also given there.

Reverting to (**4.3**), we are concerned with the conditions under which $D(0)$ is finite, non-zero. Evidently, it is a sufficient condition that both

$$\lim_{\theta\to 0}\left(\frac{1}{B}\int_0^\theta f\,\mathrm{d}\theta\right) \text{ and } \lim_{\theta\to 0}\left(B\frac{\mathrm{d}f}{\mathrm{d}\theta}\right)$$

be finite (including zero). The slightly stronger condition that these limits be *zero* ensures that $D(0)$ is equal to D_0. Evidently, the imposition on f of the conditions

$$\lim_{\theta\to 0}\left(\frac{1}{B}\int_0^\theta f\,\mathrm{d}\theta\right)=0, \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{4.6})$$

$$\lim_{\theta\to 0}\left(B\frac{\mathrm{d}f}{\mathrm{d}\theta}\right)=0, \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{4.7})$$

ensures that the inverfc term in (**4.2**) accounts completely for the finite but non-zero value of $D(0)$ in (**4.3**).

It is simply shown that (**4.6**) and (**4.7**) are equivalent to the single condition that

$$\lim_{\theta\to 0}\frac{f}{\operatorname{inverfc}\theta}=0. \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{4.8})$$

Note that since (Philip 1960)

$$\lim_{\theta\to 0}\frac{\operatorname{inverfc}\theta}{(-\log\theta)^{\frac12}}=1, \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{4.9})$$

(**4.8**) is equivalent to

$$\lim_{\theta\to 0}\frac{f}{(-\log\theta)^{\frac12}}=0. \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{4.10})$$

f is thus free to become infinite at $\theta=0$, but it must approach infinity more slowly than $(-\log\theta)^{\frac12}$. Note that it is necessary, *but not sufficient*, that $f\to\infty$ more slowly than $\theta^{-\varepsilon}$, where ε is any non-zero positive quantity.

Accordingly, there exists an exact solution of (**1.1**) subject to (**1.2**) with $D(0)=D_0$ (finite, non-zero) corresponding to every $D(\theta)$ of the form (**4.3**) for which

(i) $(2D_0^{\frac12}\operatorname{inverfc}\theta+f)$ satisfies the appropriate conditions on F set out in Section III;

(ii) f satisfies condition (**4.8**).

A further result is that, for

$$F=2D_0^{\frac12}\operatorname{inverfc}\theta\pm f \quad (D_0 \text{ finite, non-zero}), \quad \ldots\ldots \quad (\mathbf{4.11})$$

$$D=D_0\pm\tfrac12 D_0^{\frac12}\left(\frac{2}{B}\int_0^\theta f\,\mathrm{d}\theta-B\frac{\mathrm{d}f}{\mathrm{d}\theta}\right)-\tfrac12\int_0^\theta f\,\mathrm{d}\theta\cdot\frac{\mathrm{d}f}{\mathrm{d}\theta}. \quad \ldots\ldots \quad (\mathbf{4.12})$$

For the results (**4.2**), (**4.3**), the restriction on $\mathrm{d}f/\mathrm{d}\theta$ was simply

$$\mathrm{d}f/\mathrm{d}\theta<2D_0^{\frac12}/B. \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{4.13})$$

We here require the somewhat stronger restriction

$$| \mathrm{d}f/\mathrm{d}\theta | \leqslant 2D_0^{\frac{1}{2}}/B. \qquad (4.14)$$

We note in passing that condition (**3.5**) reduces to the condition on f that

$$f(1)=0. \qquad (4.15)$$

It will be observed that a very large class of functions f satisfy these various requirements, and that many of them may be found by elementary means. Some typical results are presented in Table 2. It is, of course, a simple matter to extend this table to a much greater length; but there seems little purpose in enumerating further examples.

Table 2

SOME EXACT SOLUTIONS OF (**1.1**), (**1.2**) WITH $D(0)$ FINITE NON-ZERO

No.	D	φ	Remarks
1	$D_0+\frac{1}{2}n\theta^n\left(1-\frac{\theta^n}{1+n}\right)$ $\pm\frac{1}{2}D_0^{\frac{1}{2}}\left[\frac{2\theta}{B}\left(1-\frac{\theta^n}{1+n}\right)+Bn\theta^{n-1}\right]$	$2D_0^{\frac{1}{2}}$ inverfc $\theta\pm(1-\theta^n)$	$n>0$. For minus sign $D_0\geqslant\pi n^2/16$
2	$D_0+\frac{n}{2(n+1)}[(1-\theta)^{n-1}-(1-\theta)^{2n}]$ $\pm\frac{1}{2}D_0^{\frac{1}{2}}\left[\frac{2}{(n+1)B}[1-(1-\theta)^{n+1}]\right.$ $\left.+Bn(1-\theta)^{n-1}\right]$	$2D_0^{\frac{1}{2}}$ inverfc $\theta\pm(1-\theta)^n$	$n>0$. For minus sign D_0 has lower limit
3	$D_0+\frac{1}{2}\sin^2\frac{1}{2}\pi\theta\pm\frac{1}{4}D_0^{\frac{1}{2}}\sin\frac{1}{2}\pi\theta\cdot\left[\frac{8}{\pi B}+\pi B\right]$	$2D_0^{\frac{1}{2}}$ inverfc $\theta\pm\cos\frac{1}{2}\pi\theta$	For minus sign $D_0\geqslant\pi/16$
4	$D_0+\frac{1}{2}e^{n\theta}[1-e^{n\theta}+n\theta e^n]$ $\pm\frac{1}{2}D_0^{\frac{1}{2}}\left[\frac{2}{nB}(1-e^{n\theta}+n\theta e^n)+Bne^{n\theta}\right]$	$2D_0^{\frac{1}{2}}$ inverfc $\theta\pm(e^n-e^{n\theta})$	$n>0$. For minus sign $D_0\geqslant\pi n^2e^{2n}/16$

V. Solutions of (1.1), (1.3)

Similarly, we may establish exact solutions of (**1.1**) subject to (**1.3**) at will, simply by selecting F functions satisfying the appropriate conditions stated in Section III. Typical results are given in Table 3. It will be noted that the entries 1 and 11 of the table are for $D(\theta)$ functions symmetrical about $\theta=\frac{1}{2}$. It is evident that these results follow immediately from Nos. 2 and 4 of Table 1, and that many more " symmetrical " exact solutions may be readily constructed from the results of the preceding section.

We here again encounter the difficulty which arose previously, namely, that F functions chosen without certain precautions tend to lead to zero or infinite values of D. In Section IV, this occurred only for $D(0)$; but here the trouble arises also for $D(1)$, which we write for $\lim_{\theta\to 1} D(\theta)$.

In this case we may avoid the difficulty by using a known result for diffusion in composite media governed by conditions (**1.3**) (Carslaw and Jaeger 1959, p. 87). We find that, from the present viewpoint, this result is equivalent to the following statement.

Let F be specified by (**5.1**) below:

$$\left.\begin{array}{lll} & 1\geqslant\theta>\theta', & F=-2D_1^{\frac{1}{2}}\operatorname{inverfc}\{(1-\theta)/(1-\theta')\}\,;\\ & 0\leqslant\theta<\theta', & F=2D_0^{\frac{1}{2}}\operatorname{inverfc}(\theta/\theta')\,;\\ \text{where} & & \theta'=D_1^{\frac{1}{2}}/(D_0^{\frac{1}{2}}+D_1^{\frac{1}{2}}).\end{array}\right\}\quad\ldots\ (\mathbf{5.1})$$

TABLE 3

SOME SIMPLE CASES OF EXACT SOLUTIONS OF (**1.1**), (**1.3**)

No.	D	φ	Remarks
1	$\dfrac{n}{2(n+1)}[(1-2\theta)^{n-1}-(1-2\theta)^{2n}]$	$(1-2\theta)^n$	$n=1, 3, 5, 7, \ldots$
2	$\dfrac{n\theta^n}{2(n+1)}(1-\theta^n)$	$\dfrac{1}{n+1}-\theta^n$	$n>0$
3	$\dfrac{n\theta^{-n}}{2(1-n)}(\theta^{-n}-1)$	$\theta^{-n}-\dfrac{1}{1-n}$	$0<n<1$
4	$\dfrac{n(1-\theta)^n}{2(n+1)}[1-(1-\theta)^n]$	$(1-\theta)^n-\dfrac{1}{n+1}$	$n>0$
5	$\dfrac{n(1-\theta)^{-n}}{2(1-n)}[(1-\theta)^{-n}-1]$	$\dfrac{1}{1-n}-(1-\theta)^{-n}$	$0<n<1$
6	$\frac{1}{2}\sin\frac{1}{2}\pi\theta(\sin\frac{1}{2}\pi\theta-\theta)$	$\cos\frac{1}{2}\pi\theta-2/\pi$	
7	$\frac{1}{2}\cos\frac{1}{2}\pi\theta[\cos\frac{1}{2}\pi\theta+\theta-1]$	$2/\pi-\sin\frac{1}{2}\pi\theta$	
8	$\frac{1}{2}\left[\dfrac{1+\theta\cos^{-1}\theta-\theta}{\sqrt{(1-\theta^2)}}+1\right]$	$\cos^{-1}\theta-1$	
9	$\frac{1}{2}\left[\dfrac{(1-\theta)\cos^{-1}(1-\theta)+\theta}{\sqrt{(2\theta-\theta^2)}}\right]$	$\sin^{-1}(1-\theta)-\frac{1}{2}\pi+1$	
10	$\frac{1}{4}\theta\left[\dfrac{\theta+\sin^{-1}\theta-\frac{1}{2}\pi\theta}{\sqrt{(1-\theta^2)}}\right]$	$\sqrt{(1-\theta^2)}-\frac{1}{4}\pi$	
11	$\frac{1}{2}\sin^2\pi\theta$	$\cos\pi\theta$	
12	$\dfrac{\sin^2\pi\theta}{16}[5+2\cos\pi\theta-4\theta\operatorname{cosec}\frac{1}{2}\pi\theta]$	$\cos^3\frac{1}{2}\pi\theta-4/3\pi$	
13	$\frac{1}{2}n^2e^{n\theta}[(e^n-1)\theta-(e^{n\theta}-1)]$	$e^n-1-ne^{n\theta}$	n may be either positive or negative
14	$-\frac{1}{2}\log\theta$	$-(\log\theta+1)$	
15	$\dfrac{n}{2\theta}(-\log\theta)^{n-1}$ $\times\{\Gamma(n+1)\cdot(1-\theta)-\Gamma(-\log\theta, n+1)\}$	$(-\log\theta)^n-\Gamma(n+1)$	$n>0$

Then the corresponding D function is :

$$\left.\begin{array}{ll} 1\geqslant\theta>\theta', & D=\text{constant}=D_1\,; \\ 0\leqslant\theta<\theta', & D=\text{constant}=D_0. \end{array}\right\} \qquad \ldots\ldots\ldots (5.2)$$

We therefore put F into the form (D_1, D_0 finite, non-zero)

$$\left.\begin{array}{ll} 1\geqslant\theta>\theta', & F=-2D_1^{\frac{1}{2}}\,\text{inverfc}\,\{(1-\theta)/(1-\theta')\}+f(\theta)\,; \\ 0\leqslant\theta<\theta', & F=2D_0^{\frac{1}{2}}\,\text{inverfc}\,(\theta/\theta')+f(\theta). \end{array}\right\} \qquad \ldots (5.3)$$

Using (**3.4**), we obtain the corresponding $D(\theta)$ for diffusion subject to conditions (**1.3**) :

$$\left.\begin{array}{ll} 1\geqslant\theta>\theta', & D=D_1+\tfrac{1}{2}D_1^{\frac{1}{2}}\left[\dfrac{2}{(1-\theta')B\{(1-\theta)/(1-\theta')\}}\displaystyle\int_0^\theta f\,\mathrm{d}\theta-(1-\theta')B\left(\dfrac{1-\theta}{1-\theta'}\right)\dfrac{\mathrm{d}f}{\mathrm{d}\theta}\right] \\ & \qquad -\tfrac{1}{2}\dfrac{\mathrm{d}f}{\mathrm{d}\theta}\cdot\displaystyle\int_0^\theta f\,\mathrm{d}\theta\,; \\ 0\leqslant\theta<\theta', & D=D_0+\tfrac{1}{2}D_0^{\frac{1}{2}}\left[2/\theta'B(\theta/\theta')\displaystyle\int_0^\theta f\,\mathrm{d}\theta-\theta'B(\theta/\theta')\dfrac{\mathrm{d}f}{\mathrm{d}\theta}\right] \\ & \qquad -\tfrac{1}{2}\dfrac{\mathrm{d}f}{\mathrm{d}\theta}\cdot\displaystyle\int_0^\theta f\,\mathrm{d}\theta. \end{array}\right\} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (5.4)$$

The condition that $D(0)=D_0$ reduces to

$$\lim_{\theta\to 0}\frac{f}{\text{inverfc}\,(\theta/\theta')}=0, \qquad \ldots\ldots\ldots\ldots (5.5)$$

entirely analogous to (**4.8**) in the previous Section. We have similarly that the condition that $D(1)=D_1$ is that

$$\lim_{\theta\to 1}\frac{f}{\text{inverfc}\,\{(1-\theta)/(1-\theta')\}}=0. \qquad \ldots\ldots\ldots (5.6)$$

It follows that there exists an exact solution of (**1.1**) subject to (**1.3**) with $D(0)=D_0$, $D(1)=D_1$ (both finite, non-zero) corresponding to every $D(\theta)$ of the form (**5.4**) for which

(i) F of the form (**5.3**) satisfies the appropriate conditions set out in Section III ;

(ii) f satisfies conditions (**5.5**) and (**5.6**).

It will be noted that, in this case, when $\mathrm{d}f/\mathrm{d}\theta$ is continuous at $\theta=\theta'$, a discontinuity of magnitude D_1-D_0 occurs in $D(\theta)$ at this point. The condition that $\mathrm{d}F/\mathrm{d}\theta$, and therefore $D(\theta)$, is continuous at $\theta=\theta'$ is that

$$\frac{\mathrm{d}f}{\mathrm{d}\theta}(\theta'+)-\frac{\mathrm{d}f}{\mathrm{d}\theta}(\theta'-)=\left(\frac{\pi}{D_0D_1}\right)^{\frac{1}{2}}(D_1^{\frac{1}{2}}+D_0^{\frac{1}{2}})(D_1-D_0). \qquad \ldots (5.7)$$

That is, a discontinuity in $\mathrm{d}f/\mathrm{d}\theta$, defined in (**5.7**), is necessary if a discontinuity in $D(\theta)$ is to be avoided at $\theta=\theta'$. It is seen that usually f will be required to satisfy this further condition.

It is therefore helpful to replace f by $g+f'$, where the discontinuity in $\mathrm{d}f/\mathrm{d}\theta$ at $\theta=\theta'$ is contained in $\mathrm{d}g/\mathrm{d}\theta$, with $\mathrm{d}f'/\mathrm{d}\theta$ continuous at this point. Condition (**3.6**) on F reduces here to the requirement that

$$\int_0^1 f\mathrm{d}\theta=0,$$

so that it is convenient to take g such that

$$\int_0^1 g\mathrm{d}\theta=0. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{5.8})$$

We then have as a condition on f' that

$$\int_0^1 f'\mathrm{d}\theta=0. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{5.9})$$

The family of linear g functions satisfying (**5.8**) is

$$\left.\begin{array}{ll} 1\geqslant\theta>\theta', & g=a+b(\theta-\theta')\,; \\ 0\leqslant\theta<\theta', & g=a+(b-c)(\theta-\theta'). \end{array}\right\} \quad \ldots\ldots\ldots \quad (\mathbf{5.10})$$

Here a is an arbitrary constant,

$$b=(D_1^{\frac{1}{2}}+D_0^{\frac{1}{2}})[2a/(D_1^{\frac{1}{2}}-D_0^{\frac{1}{2}})+(\pi D_1/D_0)^{\frac{1}{2}}],$$

and

$$c=(\pi/D_0D_1)^{\frac{1}{2}}(D_1^{\frac{1}{2}}+D_0^{\frac{1}{2}})(D_1-D_0).$$

We recall that g and f' must be such that $\mathrm{d}F/\mathrm{d}\theta\leqslant 0$.

It is evident that, despite these various restrictions, there still remains a wide range of permissible f functions. Figures 1 and 2 represent some simple exact solutions for the case $D_0=1$, $D_1=9$, $f'=0$, and several values of the parameter a.

We have here treated the case with both D_0 and D_1 finite non-zero. Evidently, when only *one* of these quantities is required to be finite non-zero, the conditions on F are weaker and the problem is correspondingly simpler.

VI. Practical Use of the Method

The typical mathematical problem arising in concentration-dependent diffusion is to compute $\varphi(\theta)$ when $D(\theta)$ is known (frequently determined experimentally). Effective use of the method developed here depends then on the ability to match any given $D(\theta)$ function with a member of the family of $D(\theta)$ functions yielding exact solutions.

This may be done, though rather brutally, by introducing (for the case of diffusion subject to (**1.2**)) an F function of the form

$$F=2D_0\,\mathrm{inverfc}\,\theta+\sum_1^n a_r(1-\theta)^r. \quad \ldots\ldots\ldots\ldots \quad (\mathbf{6.1})$$

The corresponding $D(\theta)$ function follows from (**3.4**). We may now match this computed $D(\theta)$ with its $(n+1)$ adjustable parameters, D_0 and a_r, at $(n+1)$

points of the given $D(\theta)$ curve. The problem reduces to solving $(n+1)$ simultaneous equations for $(n+1)$ unknowns. This is a possible, but not particularly attractive, manner of securing the solution. A similar method may be used for solving (**1.1**) subject to (**1.3**).

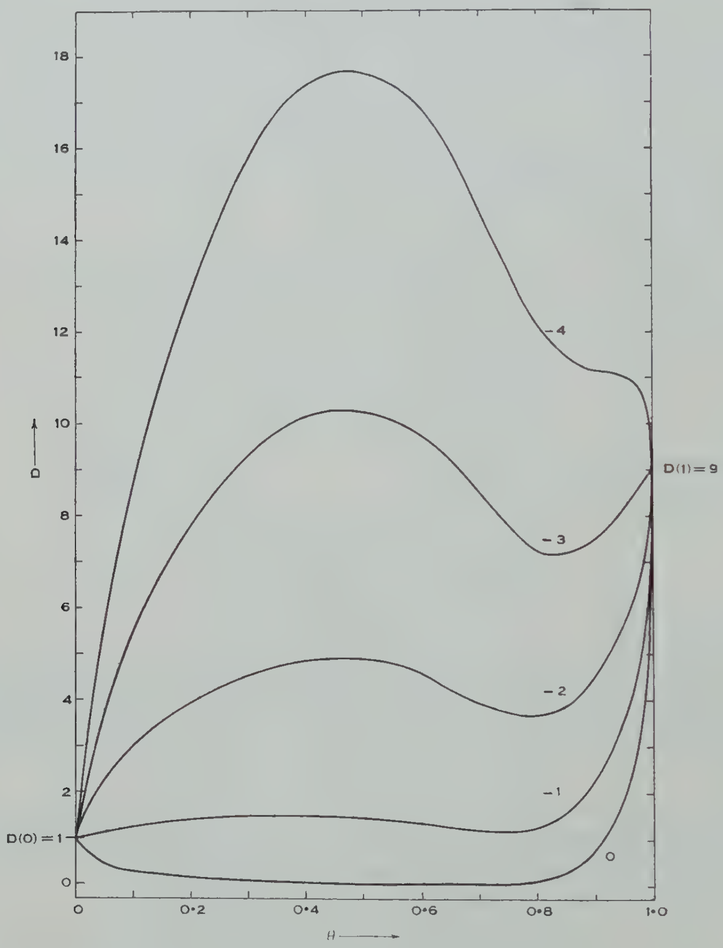


Fig. 1.—Parametric family of $D(\theta)$ functions defined as follows:

$1 \geqslant \theta > \frac{3}{4}$, $D = 9 + (1-\theta)[(2a+6\sqrt{\pi})\theta - 3\sqrt{\pi}]\{2a+6\sqrt{\pi} - 12/B(4-4\theta)\} - \frac{3}{4}(2a+6\sqrt{\pi})B(4-4\theta)$;

$0 \leqslant \theta < \frac{3}{4}$, $D = 1 - \theta[(2a+\frac{2}{3}\sqrt{\pi})\theta - (2a+\sqrt{\pi})]\{2a+\frac{2}{3}\sqrt{\pi} - 4/3B(4\theta/3)\} - \frac{3}{4}(2a+\frac{2}{3}\sqrt{\pi})B(4\theta/3)$.

Numbers on the curves denote values of a.

It appears to the author that (both in the case of (**1.1**), (**1.2**) and of (**1.1**), (**1.3**)) it is preferable to keep the number of disposable constants in the F function down to perhaps two or three, and to provide the desired range of shapes of $D(\theta)$ function by making use of the variety of the possible functional forms which F is free to assume. This requires the construction of a "library" in which are arrayed various F functions together with the corresponding D

functions. It should then be possible, by inspection, to select an appropriate functional form of F, and the problem then reduces to matching the given D function by solving for the small number of disposable parameters occurring in the F function. It is beyond the scope of the present paper to provide such a library. Its construction and presentation remain a task for the future.

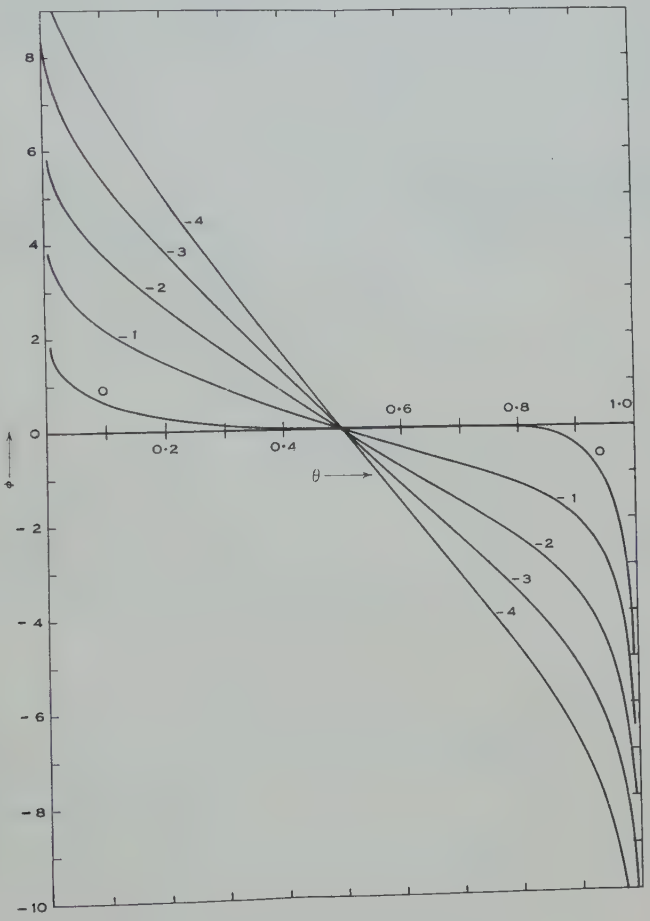


Fig. 2.—Parametric family of $\varphi(\theta)$ functions defined as follows:

$1 \geqslant \theta > \frac{3}{4}$, $\varphi = -6 \operatorname{inverfc}(4-4\theta) + a + 4(a + 3\sqrt{\pi})(\theta - \frac{3}{4})$;

$0 \leqslant \theta < \frac{3}{4}$, $\varphi = 2 \operatorname{inverfc} 4\theta/3 + a + 4(a + \frac{1}{3}\sqrt{\pi})(\theta - \frac{3}{4})$.

Numbers on the curves denote values of a. These curves represent exact solutions of (**1.1**), (**1.3**) corresponding to the $D(\theta)$ functions of Figure 1.

VII. DISCUSSION

In the present study we have made use of (**4.1**) and (**5.1**), (**5.2**) to avoid difficulties due to $D(\theta)$ becoming zero or infinite at one or both ends of the θ-interval; but we have not provided a thorough investigation of such questions as the relationship (for diffusion subject to (**1.2**)) between the manner in which

$\varphi(\theta)$ behaves as $\theta \to 0$ and the behaviour of $D(\theta)$ close to $\theta = 0$. It is hoped to return to questions of this nature in a later communication (cf. Philip 1957 for an early approach).

VIII. REFERENCES

CARSLAW, H. S., and JAEGER, J. C. (1959).—" Conduction of Heat in Solids." 2nd Ed. (Clarendon Press : Oxford.)

CRANK, J. (1956).—" Mathematics of Diffusion." (Clarendon Press : Oxford.)

FUJITA, H. (1952*a*).—*Text. Res. J.* **22** : 757.

FUJITA, H. (1952*b*).—*Text. Res. J.* **22** : 823.

FUJITA, H. (1954).—*Text. Res. J.* **24** : 234.

PHILIP, J. R. (1955).—*Trans. Faraday Soc.* **51** : 885.

PHILIP, J. R. (1956).—*Aust. J. Phys.* **9** : 570.

PHILIP, J. R. (1957).—*Soil Sci.* **84** : 163.

PHILIP, J. R. (1960).—*Aust. J. Phys.* **13** : 13.

THE FUNCTION INVERFC θ

By J. R. Philip*

[*Manuscript received November* 11, 1959]

Summary

The function inverfc θ arises in certain diffusion problems when concentration is taken as an independent variable. It enters into a general method of exact solution of the concentration-dependent diffusion equation. An account is given of the properties of this function, and of its derivatives and integrals. The function

$$B(\theta)=(2/\pi^{\frac{1}{2}})\exp[-(\text{inverfc}\,\theta)^2]$$

is intimately connected with the first integral of inverfc θ and with its derivatives. Tables of inverfc θ and $B(\theta)$ are given.

I. Introduction

The solution of one-dimensional diffusion problems is usually sought in the form

concentration = explicit function of distance and time.

It has become increasingly evident, however, that there are occasions when it is simpler, and more illuminating, to seek the solution in the form

distance = explicit function of concentration and time.

In particular, the latter approach has proved fruitful when applied to concentration-dependent diffusion (Philip 1955) and when applied to problems where concentration-dependent diffusion is combined with a first-order (not necessarily linear) phenomenon (Philip 1957).

In these connexions, it was found convenient (Philip 1955) to introduce the notation " inverfc " to denote the inverse of the function

$$\text{erfc}\,x=\frac{2}{\pi^{\frac{1}{2}}}\int_x^\infty \exp(-\zeta^2)\mathrm{d}\zeta. \quad \ldots\ldots\ldots\ldots \quad (1.1)$$

Until now there has been no urgent need to examine in detail the properties of the inverfc function. However, inverfc θ and its first derivative and first integral with respect to θ enter intimately into the recently found general method of exact solution of the concentration-dependent diffusion equation (Philip 1960). This account of the properties of inverfc θ, its derivatives, and its integrals, therefore forms an essential supplement to Philip (1960). The tabulations of the functions inverfc θ and $B(\theta)$ given here will frequently be needed when the method of Philip (1960) is applied.

---

* Division of Plant Industry, C.S.I.R.O., Canberra.

The concept of inverfc as the inverse of erfc leads to definition of the function through (**1.2**):

$$\theta=\operatorname{erfc}\ (\operatorname{inverfc}\theta). \qquad \dots\dots\dots\dots\dots\dots \quad (\mathbf{1.2})$$

An alternative *de novo* definition of inverfc follows from equations (**3.4**) and (**4.1**) of Philip (1960). In this way the function may be introduced as the solution, $F=\operatorname{inverfc}\theta$, of the equation

$$\frac{\mathrm{d}F}{\mathrm{d}\theta}\int_0^\theta F\mathrm{d}\theta=-\tfrac{1}{2}, \qquad \dots\dots\dots\dots\dots\dots \quad (\mathbf{1.3})$$

subject to the conditions

$$F(1)=0\ ;\quad 0\leqslant\theta\leqslant 1,\ F\geqslant 0. \qquad \dots\dots\dots\dots \quad (\mathbf{1.4})$$

The following elementary results come directly from the known properties of erfc:

$$\operatorname{inverfc}0=+\infty\ ;\quad \operatorname{inverfc}1=0\ ;\quad \operatorname{inverfc}2=-\infty, \qquad \dots \quad (\mathbf{1.5})$$

$$\operatorname{inverfc}(2-\theta)=-\operatorname{inverfc}\theta. \qquad \dots\dots\dots\dots\dots\dots\dots \quad (\mathbf{1.6})$$

We shall deal almost exclusively with the interval in θ, $0\leqslant\theta\leqslant 1$; it is a trivial matter to extend the results to the whole interval $0\leqslant\theta\leqslant 2$ by means of (**1.6**). Note that inverfc θ is defined only within the latter interval.

Table 1

inverfc θ computed from series (**2.4**)

θ	inverfc θ Computed from Series (**2.4**)			inverfc θ Exact Value
	Three Terms	Four Terms	Five Terms	
0·6	0·3706458	0·3707451	0·3707570	0·3708072
0·7	0·2724424	0·2724557	0·2724566	0·2724627
0·8	0·1791423	0·1791431	0·1791431	0·1791434
0·9	0·0888560	0·0888560	0·0888560	0·0888560

II. Power Series for inverfc θ

We introduce the power series connected with erfc:

$$\tfrac{1}{2}\pi^{\frac{1}{2}}(1-\operatorname{erfc}x)=x-\frac{x^3}{3}+\frac{x^5}{2!5}-\frac{x^7}{3!7}+\frac{x^9}{4!9}-\dots \quad (\mathbf{2.1})$$

Putting

$$\vartheta=\tfrac{1}{2}\pi^{\frac{1}{2}}(1-\theta), \qquad \dots\dots\dots\dots\dots\dots\dots \quad (\mathbf{2.2})$$

and writing x for inverfc θ, we have

$$\vartheta=x-\frac{x^3}{3}+\frac{x^5}{2!5}-\frac{x^7}{3!7}+\frac{x^9}{4!9}-\dots \qquad \dots\dots\dots \quad (\mathbf{2.3})$$

Suppose now, that inverfc θ (i.e. x) may be expanded as a power series in ϑ. Then we may formally establish this series by equating coefficients of powers of ϑ on each side of (2.3). The result is

$$\operatorname{inverfc} \theta = \vartheta + \tfrac{1}{3}\vartheta^3 + \frac{7}{30}\vartheta^5 + \frac{127}{630}\vartheta^7 + \frac{4369}{22680}\vartheta^9 + \ldots \qquad \ldots (2.4)$$

No simple general expression for the coefficients is apparent. It will be shown later in Section IV that (2.4) may be derived directly from the Taylor expansion of inverfc θ about $\theta = 1$. The series of (2.3) is uniformly convergent. Presumably the series of (2.4) converges for $|\vartheta| < \tfrac{1}{2}\pi^{\frac{1}{2}}$. It provides a useful means of calculating inverfc θ in the neighbourhood of $\theta = 1$. Table 1 gives a comparison of the exact value of inverfc θ with that computed from the first few terms of series (2.4).

III. Asymptotic Forms of inverfc θ, θ Small

For large values of x (=inverfc θ), we have the well-known asymptotic result:

$$\theta \approx \frac{\exp(-x^2)}{\pi^{\frac{1}{2}} x}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.1)$$

(3.1) is equivalent to

$$x^2 \approx -\log \theta - \tfrac{1}{2} \log \pi x^2,$$

which has the continued logarithmic form

$$x^2 \approx -\log \theta - \tfrac{1}{2} \log [\pi(-\log \theta - \tfrac{1}{2} \log [\,.\ .\ .\ .$$

Accordingly we have the approximation for θ small:

$$\operatorname{inverfc} \theta = \{-\log \theta - \tfrac{1}{2} \log [\pi(-\log \theta - \tfrac{1}{2} \log [\,.\ .\ .\}^{\frac{1}{2}}. \qquad \ldots (3.2)$$

As far as the author knows, no formal study has been made of the convergence of continued logarithms. The convergence of (3.2) is rapid for θ small. See Table 2. In this table the symbol S_n denotes the nth member of the sequence formed by terminating the repeated logarithm at successive log θ's. Thus,

$$S_1 = (-\log \theta)^{\frac{1}{2}};$$
$$S_2 = (-\log \theta - \tfrac{1}{2} \log [\pi(-\log \theta)])^{\frac{1}{2}};$$
$$S_3 = (-\log \theta - \tfrac{1}{2} \log [\pi(-\log \theta - \tfrac{1}{2} \log \{\pi(-\log \theta)\})])^{\frac{1}{2}};$$

and so on.

It is evident that the limit to the accuracy of using (3.2) for θ small is set by the limited accuracy of (3.1) rather than by the rate of convergence of the sequence S_n. Note that S_2 proves a better approximation to inverfc θ than do the higher members of the sequence.

TABLE 2
inverfc θ COMPUTED FROM (**3.2**)

θ	S_1	S_2	S_3	S_4	inverfc θ Exact Value
10^{-6}	3·7169	3·4540	3·4646	3·4642	3·4589
10^{-4}	3·0349	2·7437	2·7620	2·7608	2·7509
10^{-2}	2·1460	1·8081	1·8549	1·8480	1·8214

Pollack (1956) has established an inequality which leads to the following improved approximation

$$\theta\approx\frac{2\exp(-x^2)}{\pi^{\frac12}\{x+\sqrt{(x^2+4/\pi)}\}}. \qquad \ldots\ldots\ldots\ldots\ (\mathbf{3.3})$$

This yields the better approximation

$$\operatorname{inverfc}\theta=\{-\log\theta-\sinh^{-1}\sqrt{\tfrac14\pi(-\log\theta-\sinh^{-1}\sqrt{[\tfrac14\pi(-\log\theta\ .\ .\ .}}\}^{\frac12}. \qquad \ldots\ldots\ldots\ (\mathbf{3.4})$$

(**3.4**) is more accurate than (**3.2**), and converges at about the same rate ; but it involves $\sinh^{-1}$, which is scarcely any simpler than inverfc.

The following result follows from (**3.2**) or (**3.3**)

$$\lim_{\theta\to0}\frac{\operatorname{inverfc}\theta}{(-\log\theta)^{\frac12}}=1.$$

IV. THE DERIVATIVES AND INTEGRALS OF inverfc θ

Differentiating equation (**1.2**), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}\theta}(\operatorname{inverfc}\theta)=-\tfrac12\pi^{\frac12}\exp[(\operatorname{inverfc}\theta)^2]. \qquad \ldots\ldots\ldots\ldots\ (\mathbf{4.1})$$

It is convenient to introduce the function

$$B(\theta)=\frac{2}{\pi^{\frac12}}\exp[-(\operatorname{inverfc}\theta)^2]. \qquad \ldots\ldots\ldots\ldots\ (\mathbf{4.2})$$

We then have

$$\frac{\mathrm{d}}{\mathrm{d}\theta}(\operatorname{inverfc}\theta)=-\frac{1}{B}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ (\mathbf{4.3})$$

In addition,

$$\frac{\mathrm{d}B}{\mathrm{d}\theta}=2\operatorname{inverfc}\theta, \qquad \ldots\ldots\ldots\ldots\ldots\ (\mathbf{4.4})$$

and, in general ($m\neq0$, $n\neq0$),

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left[\frac{(\operatorname{inverfc}\theta)^m}{B^n}\right]=-\frac{1}{B^{n+1}}[2n(\operatorname{inverfc}\theta)^{m+1}+m(\operatorname{inverfc}\theta)^{m-1}]. \qquad \ldots\ (\mathbf{4.5})$$

This result is useful for generating the higher derivatives[*] of inverfc θ. We have, for example,

$$\frac{d^2}{d\theta^2}(\text{inverfc }\theta)=\frac{1}{B^2}(2\text{ inverfc }\theta),$$

$$\frac{d^3}{d\theta^3}(\text{inverfc }\theta)=-\frac{1}{B^3}\{8(\text{inverfc }\theta)^2+2\},$$

$$\frac{d^4}{d\theta^4}(\text{inverfc }\theta)=\frac{1}{B^4}\{48(\text{inverfc }\theta)^3+28\text{ inverfc }\theta\},$$

$$\frac{d^5}{d\theta^5}(\text{inverfc }\theta)=-\frac{1}{B^5}\{384(\text{inverfc }\theta)^4+368(\text{inverfc }\theta)^2+28\},$$

$$\frac{d^6}{d\theta^6}(\text{inverfc }\theta)=\frac{1}{B^6}\{3840(\text{inverfc }\theta)^5+5216(\text{inverfc }\theta)^3+1016\text{ inverfc }\theta\},$$

$$\frac{d^7}{d\theta^7}(\text{inverfc }\theta)=-\frac{1}{B^7}\{46{,}080(\text{inverfc }\theta)^6+81{,}792(\text{inverfc }\theta)^4$$
$$+27{,}840(\text{inverfc }\theta)^2+1016\}.$$

It is evident that, *for* n *odd*,

$$\frac{d^n}{d\theta^n}(\text{inverfc }\theta)=-\frac{1}{B^n}(a_{n-1}(\text{inverfc }\theta)^{n-1}+a_{n-3}(\text{inverfc }\theta)^{n-3}+\ldots+a_0),$$

and, *for* n *even*,

$$\frac{d^n}{d\theta^n}(\text{inverfc }\theta)=\frac{1}{B^n}\{a_{n-1}(\text{inverfc }\theta)^{n-1}+a_{n-3}(\text{inverfc }\theta)^{n-3}$$
$$+\ldots+a_1\text{ inverfc }\theta\}.$$

In both cases the coefficients a_{n-1}, etc. are all positive.

Now $2/\pi^{\frac{1}{2}}\geqslant B\geqslant 0$ throughout the interval $0\leqslant\theta\leqslant 2$, whilst inverfc θ is positive in $0\leqslant\theta<1$, zero at $\theta=1$, and negative in $1<\theta\leqslant 2$. It therefore follows that, in the interval $0\leqslant\theta<1$,

$\frac{d^n}{d\theta^n}$ (inverfc θ) is positive if n is even, negative if n is odd;

at $\theta=1$,

$\frac{d^n}{d\theta^n}$ (inverfc θ) is zero if n is even, negative if n is odd;

in the interval $1<\theta\leqslant 2$,

$\frac{d^n}{d\theta^n}$ (inverfc θ) is negative whether n is even or odd.

---

[*] I am indebted to the referee for remarks which suggest the following, more elegant, treatment of the higher derivatives of inverfc θ.

$F=$inverfc θ satisfies the equation

$$d^2F/d\theta^2=2F(dF/d\theta)^2. \qquad \ldots\ldots \text{(A)}$$

This may be established by differentiating (**1.3**) with respect to θ. Now, if $P_n(F)$ denotes a polynomial in F, and

$$d^nF/d\theta^n=P_n(F)(dF/d\theta)^n. \qquad \ldots\ldots \text{(B)}$$

it follows by differentiation and use of (**A**) that

$$d^{n+1}F/d\theta^{n+1}=P_{n+1}(F)(dF/d\theta)^{n+1},$$

where

$$P_{n+1}(F)=(2nF+d/dF)P_n(F). \qquad \ldots\ldots \text{(C)}$$

Now (**B**) is true for $n=1$, and $P_1(F)=1$. Therefore, $P_n(F)$ for all $n>1$ follows at once from (**C**), giving a simple means of determining the higher derivatives of inverfc θ.

We have the particular results :

$$(-1)^n \frac{\mathrm{d}^n}{\mathrm{d}\theta^n} (\text{inverfc } 0) = \infty,$$

$$\frac{\mathrm{d}^n}{\mathrm{d}\theta^n} (\text{inverfc } 2) = -\infty.$$

The values of the first nine derivatives of inverfc θ at $\theta=1$ are

$$-\tfrac{1}{2}\pi^{\frac{1}{2}},\ 0,\ -2(\tfrac{1}{2}\pi^{\frac{1}{2}})^3,\ 0,\ -28(\tfrac{1}{2}\pi^{\frac{1}{2}})^5,\ 0,\ -1016(\tfrac{1}{2}\pi^{\frac{1}{2}})^7,\ 0,\ -69{,}904(\tfrac{1}{2}\pi^{\frac{1}{2}})^9.$$

It follows from these results that equation (**2.4**) may be established by applying Taylor's theorem to the right-hand side of the identity

$$\text{inverfc } \theta = \text{inverfc } [1-(1-\theta)].$$

We also note that it follows from (**4.4**) that

$$\frac{\mathrm{d}^n B}{\mathrm{d}\theta^n} = 2 \frac{\mathrm{d}^{n-1}}{\mathrm{d}\theta^{n-1}} (\text{inverfc } \theta). \qquad \ldots\ldots\ldots\ldots \quad (\mathbf{4.6})$$

It is readily established by integration by parts, or by use of (**4.1**) in (**1.3**), or by integrating (**4.4**), that

$$\int_0^\theta \text{inverfc } \theta \, \mathrm{d}\theta = \frac{1}{\pi^{\frac{1}{2}}} \exp\,(-\text{inverfc } \theta)^2 = \tfrac{1}{2}B. \qquad \ldots \quad (\mathbf{4.7})$$

A further integration yields

$$\int_0^\theta \int_0^\theta \text{inverfc } \theta \, \mathrm{d}\theta \mathrm{d}\theta = \frac{1}{\sqrt{(2\pi)}} \,\text{erfc}\,(\sqrt{2} \text{ inverfc } \theta).$$

This and the higher integrals of inverfc θ do not appear to be of significance in the present developments.

V. The Function $B(\theta)$

We have seen that $B(\theta)$ is simply related to the first derivative, and to the first integral, of inverfc θ. For this reason it proves of primary importance in the development of the general method of exact solution of the concentration-dependent diffusion equation (Philip 1960).

We note from (**1.6**) and (**4.2**) that

$$B(2-\theta) = B(\theta). \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{5.1})$$

We have already remarked on the simple relation between derivatives of inverfc θ and those of B. It follows that the Taylor expansion of $B(\theta)$ about $\theta=1$ yields as the power series for B

$$B(\theta) = \frac{2}{\pi^{\frac{1}{2}}} \left[1 - \vartheta^2 - \frac{\vartheta^4}{6} - \frac{7}{90}\vartheta^6 - \frac{127}{2520}\vartheta^8 - \frac{4369}{113400}\vartheta^{10} - \ldots\right], \quad \ldots \quad (\mathbf{5.2})$$

where ϑ is again defined by (**2.2**). Presumably the series of (**5.2**) converges for $|\vartheta| \leqslant \frac{1}{2}\pi^{\frac{1}{2}}$. It enables B to be calculated readily in the neighbourhood of $\theta=1$.

The behaviour of B near $\theta=0$ is of interest. Approximation (**3.1**) applied in (**4.2**) yields

$$B(\theta)\approx 2\theta(-\log\theta)^{\frac{1}{2}}, \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{5.3})$$

and it may be shown that

$$\lim_{\theta\to 0}\frac{B(\theta)}{2\theta\,(-\log\theta)^{\frac{1}{2}}}=1.$$

It follows that, as $\theta\to 0$, $B\to 0$ more rapidly* than does $\theta^{1-\varepsilon}$, where ε is any non-zero positive quantity, and more slowly than does θ.

VI. TABLES OF inverfc θ AND $B(\theta)$

The only existing table of inverfc θ known to the author is in Fowle (1921, p. 60). The column of the table headed v/c gives θ in the present notation, and that headed $2q$ gives values of 2 inverfc θ to four or five significant figures. In connexion with Philip (1960) it is helpful to have a table of inverfc θ readily available. In the course of constructing the table of $B(\theta)$, it was a simple matter to develop a new and more accurate table of inverfc θ. Details are given below, and the resulting tabulation is presented in Table 3. No graph of inverfc θ is given, since the shape of erfc x is well known.

TABLE 3

THE FUNCTIONS inverfc θ AND $B(\theta)$

θ	inverfc θ	$B(\theta)$	θ	inverfc θ	$B(\theta)$
0	∞	0	0·40	0·595 116 1	0·791 851 9
10^{-6}	3·458 911	0·000 007 177 8	0·45	0·534 159 1	0·848 280 6
10^{-5}	3·122 587	0·000 065 742	0·50	0·476 936 3	0·898 807 9
10^{-4}	2·750 936	0·000 583 27	0·55	0·422 680 2	0·943 766 6
10^{-3}	2·326 754	0·005 026 6	0·60	0·370 807 2	0·983 423 2
10^{-2}	1·821 386	0·040 898 3	0·65	0·320 858 3	1·017 992 1
0·05	1·385 904	0·165 219 5	0·70	0·272 462 7	1·047 646 6
0·10	1·163 087	0·291 711 6	0·75	0·225 312 1	1·072 526 1
0·15	1·017 902	0·400 379 2	0·80	0·179 143 4	1·092 741 7
0·20	0·906 193 8	0·496 384 2	0·85	0·133 726 9	1·108 379 8
0·25	0·813 419 8	0·582 241 7	0·90	0·088 856 0	1·119 505 3
0·30	0·732 869 1	0·659 472 6	0·95	0·044 341 3	1·126 162 8
0·35	0·660 854 4	0·729 098 6	1·00	0	1·128 379 2

No table of $B(\theta)$ is known to the author. The tabulation of this function, which is also given in Table 3, was constructed with the aid of National Bureau of Standards (1954) tables by methods described below. Figure 1 gives the plot of $B(\theta)$.

Table of inverfc θ. The table of inverfc θ was constructed from the National Bureau of Standards (1954) tables of erf x by a process of inverse interpolation.

---

* Suppose $\lim_{\theta\to\theta_0} P(\theta)=0$; $\lim_{\theta\to\theta_0} Q(\theta)=0$. Then we say that, as $\theta\to\theta_0$, $P\to 0$ more rapidly than does Q, provided that $\lim_{\theta\to\theta_0}\frac{P}{Q}(\theta)=0$.

" Linear " interpolation of the type suggested in the introduction to those tables proved sufficiently accurate to ensure that errors in the final place given would not exceed unity. Twenty comparisons with Fowle's table were possible ; in every case the final place of Fowle's tabulation of $2q$ (i.e. 2 inverfc θ) was confirmed.

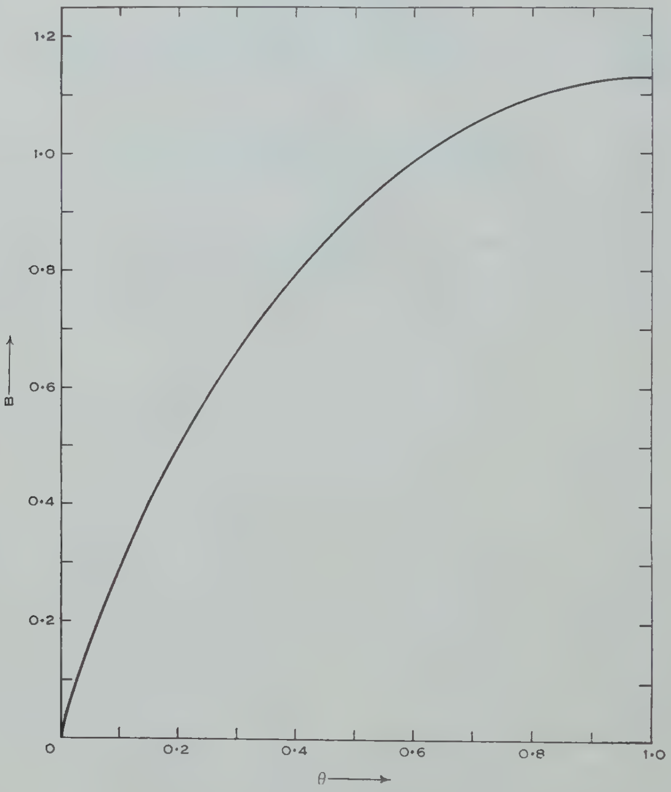


Fig. 1.—The function $B(\theta)$.

Table of $B(\theta)$. Once inverfc θ was computed, it was a simple matter to calculate $B(\theta)$ by linear interpolation (again of the type suggested in the introduction) in the tables of the derivative of erf x in the National Bureau of Standards tables. It was established that this process would not yield errors greater than unity in the final places shown in the table.

Most of each table was checked by differencing.

VII. References

Fowle, F. E. (1921).—" Smithsonian Physical Tables." 7th Ed. (Smithsonian Institution : Washington.)

National Bureau of Standards (1954).—" Tables of the Error Function and its Derivative." Applied Maths. Series 41. (U.S. Govt. Printing Office : Washington.)

Philip, J. R. (1955).—*Trans. Faraday Soc.* **51** : 885.

Philip, J. R. (1957).—*Aust. J. Phys.* **10** : 29.

Philip, J. R. (1960).—*Aust. J. Phys.* **13** : 1.

Pollack, H. O. (1956).—*Rep. Statist. Appl. Res. Un. Jap. Sci. Engrs.* **4** : 40.

THE DISTRIBUTION OF IONS FORMED BY ATTACHMENT OF ELECTRONS MOVING IN A STEADY STATE OF MOTION THROUGH A GAS

By C. A. Hurst* and L. G. H. Huxley†

[*Manuscript received December* 3, 1959]

Summary

The distribution of ions formed by attachment of electrons diffusing through a gas is solved exactly, and the results compared with an approximate calculation given earlier by Huxley. The corrections to the approximate results are inside the present experimental error, and so confirm the satisfactory agreement with experiment already obtained.

I. Introduction

The distribution of electrons and ions in a stream from a pole source and drifting and diffusing in a gas when the processes of ionization by collision and attachment to molecules are active, has been discussed in another paper (Huxley 1959), where an exact solution is given for the distribution of the electrons and an approximate solution for that of the ions. It was apparent that under conditions where the mean energy of agitation of an electron was as great as 20 times that of a gas molecule, or greater, the approximation to the exact solution was close and that the theory could provide the basis of an accurate method for measuring the coefficients of attachment of electrons as a function of mean energy of agitation in a given gas.

This expectation has been justified by the application of the method to oxygen in which self-consistent and accurate measurements of attachment coefficients were made (Huxley, Crompton, and Bagot 1959).

It is important, however, to extend the measurements to the range of small energies of agitation of the electrons from approximate thermal equilibrium to about 20 times this value, and for this purpose an exact solution of the problem is required. This solution is given in what follows.

II. Distribution of Electrons

Consider a pole source of electrons placed at the origin of coordinates in a gas and emitting electrons at constant rate of S per second. Let W be the drift speed of electrons through the gas in a constant and uniform electric field E and let D be the coefficient of diffusion of the electrons. Suppose that the processes of ionization by collision and electron attachment are both operative with coefficients α_i and α_a respectively. If the positive direction of the axis Oz

---

* Department of Mathematical Physics, University of Adelaide.

† Department of Physics, University of Adelaide.

be chosen to be that of W, the differential equation satisfied by concentration n of electrons in the stream is (Huxley 1959)

$$\nabla^2 n = 2\lambda \partial n/\partial z + 2\lambda\alpha n, \qquad \dots\dots\dots\dots\dots\dots \quad (1)$$

where

$$2\lambda = W/D\,; \quad \alpha = \alpha_a - \alpha_i.$$

The solution of equation (**1**) appropriate to the pole source at the origin is (Huxley 1959)

$$n(x,y,z) = (S/4\pi Dr)\exp(\lambda z - \mu r), \qquad \dots\dots\dots\dots \quad (2)$$

where

$$\mu^2 = \lambda^2 + 2\lambda\alpha\,; \quad r^2 = x^2 + y^2 + z^2.$$

The distribution n' due to a pole source of strength $S \exp(2\lambda h)$ placed at the point $(0,0,2h)$ is

$$\begin{aligned} n'(x,y,z) &= \{S \exp(2\lambda h)/4\pi Dr'\} \exp\{\lambda(z-2h) - \mu r'\} \\ &= (S/4\pi Dr')\exp(\lambda z - \mu r'), \qquad \dots\dots\dots\dots\dots\dots \quad (3) \end{aligned}$$

where

$$r' = \{x^2 + y^2 + (z-2h)^2\}^{\frac{1}{2}}.$$

It follows from equations (**2**) and (**3**) that the solution of equation (**1**) that represents the distribution $n(x,y,z)$ in a stream of electrons emitted at the rate S from a pole source at the origin and drifting and diffusing to a metal electrode coinciding with the plane $z=h$, over which n is now everywhere equal to zero, is (Huxley 1959)

$$n = (S/4\pi D)\exp(\lambda z)\cdot\{\exp(-\mu r)/r - \exp(-\mu r')/r'\}, \quad z \leqslant h, \qquad \dots\dots\dots\dots\dots\dots \quad (4)$$

since $r = r'$ over the plane $z = h$.

III. DISTRIBUTION OF IONS

Consider the distribution of negative ions formed by attachment of electrons to molecules in the stream of electrons from the isolated pole source represented by equation (**2**). Let $N(x,y,z)$ be the concentration of the negative ions. The differential equation satisfied by N is,

$$\nabla^2 N = 2\lambda_1 \partial N/\partial z - (\alpha_a W_1/D_1)n, \qquad \dots\dots\dots\dots \quad (5)$$

where

$$2\lambda_1 = W_1/D_1,$$

and W_1 and D_1 are the drift speed and the coefficient of diffusion respectively of the negative ions. The ratio λ/λ_1 is approximately (or, with a Maxwellian distribution of agitational speeds, exactly) equal to the ratio of the energies of agitation of ions and electrons respectively.

It is required to find the particular solution of equation (**5**) given that n is a solution of equation (**1**).

Put

$$P=C\exp(-kz)\int_{-\infty}^{z}\exp(kz')\cdot n\,dz', \qquad \ldots\ldots\ldots\ldots \quad (6)$$

in which C and k are constants.

It follows that

and

$$\left.\begin{aligned}\partial P/\partial z&=-kP+Cn,\\ \partial^2P/\partial z^2&=k^2P-Ckn+C\partial n/\partial z.\end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots \quad (7)$$

Also, consider

$$I=C\exp(-kz)\int_{-\infty}^{z}\exp(kz')\frac{\partial^2 n}{\partial z'^2}dz', \qquad \ldots\ldots\ldots\ldots \quad (8)$$

and let $\exp(kz')\partial n/\partial z$ and $\exp(kz')n$ both approach zero as z' tends to $-\infty$. After two successive integrations by parts and comparison with equation (7) it can be seen that

$$I=\partial^2P/\partial z^2,$$

and also

$$\nabla^2P=C\exp(-kz)\int_{-\infty}^{z}\exp(kz')\nabla'^2n\,dz'. \qquad \ldots\ldots\ldots\ldots \quad (9)$$

To solve equation (5) substitute the value $N=P$ in equation (5) and use equations (7), (8), (9), and (1), to find

$$C\int_{-\infty}^{z}\exp(kz')\{2\lambda\partial n/\partial z'+2\lambda\alpha n+2\lambda_1kn\}dz'+(\alpha_aW/D_1-2\lambda_1C)n\exp(kz)=0.$$

On differentiating this equation with respect to z and collecting terms, it reduces to

$$(2\lambda\alpha C+\alpha_aWk/D_1)n+\{\alpha_aW/D_1-2(\lambda_1-\lambda)C\}\partial n/\partial z=0,$$

whence

$$\left.\begin{aligned}C&=\frac{\alpha_aW}{2(\lambda_1-\lambda)D_1}=\frac{\lambda\alpha_a}{(\lambda_1-\lambda)}\,\frac{D}{D_1},\\ k&=-\lambda\alpha/(\lambda_1-\lambda).\end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots \quad (10)$$

Thus

$$N=\frac{\lambda\alpha_a}{(\lambda_1-\lambda)}\,\frac{D}{D_1}\exp\left(\frac{\lambda\alpha}{\lambda_1-\lambda}z\right)\int_{-\infty}^{z}\exp\left(-\frac{\lambda\alpha}{\lambda_1-\lambda}z'\right)n\,dz'. \qquad \ldots \quad (11)$$

When the solution of equation (1), $n=S\exp(\lambda z-\mu r)/4\pi Dr$, is adopted, then

$$N=\frac{S\alpha_a\lambda}{4\pi D_1(\lambda_1-\lambda)}\exp\left(\frac{\lambda\alpha}{\lambda_1-\lambda}z\right)\int_{-\infty}^{z}\exp\left\{\left(1-\frac{\alpha}{\lambda_1-\lambda}\right)\lambda z'-\mu r'\right\}dz'/r'.$$

$$\ldots\ldots\ldots\ldots \quad (12)$$

The distribution of electrons from a source at the origin and its image at $(0,0,2h)$ is given in equation (**4**) and the corresponding distribution of ions is

$$N=\frac{S\alpha_a\lambda}{4\pi D_1(\lambda_1-\lambda)}\exp\left(\frac{\lambda\alpha}{\lambda_1-\lambda}z\right)\int_{-\infty}^{z}\exp\left(1-\frac{\alpha}{\lambda_1-\lambda}\right)\lambda z'$$
$$\times\left\{\frac{\exp(-\mu r')}{r'}-\frac{\exp(-\mu r'')}{r''}\right\}\mathrm{d}z', \qquad z<h, \qquad \ldots\ldots\ldots\ldots (\mathbf{13})$$

where $r'=(\rho^2+z'^2)^{\frac{1}{2}}$, $r''=\{\rho^2+(z-2h)^2\}^{\frac{1}{2}}$, and $\rho^2=x^2+y^2$.

The current density J at a point (ρ,h) on the receiving electrode is $J=(\varepsilon NW_1)_{z-h}$ and when $\lambda_1\gg\lambda$ this expression becomes essentially the same as that derived by Huxley (1959, equation (**25**)).

An alternative solution of equation (**5**), namely,

$$N=-\frac{S\lambda\alpha_a}{4\pi D_1(\lambda_1-\lambda)}\exp\left(\frac{\lambda\alpha}{\lambda_1-\lambda}z\right)\int_{z}^{\infty}\exp\left(1-\frac{\alpha}{\lambda_1-\lambda}\right)\lambda z'$$
$$\times\left\{\frac{\exp(-\mu r')}{r'}-\frac{\exp(\mu r'')}{r''}\right\}\mathrm{d}z'$$

is inadmissible since in it N does not vanish at $z=-\infty$.

Equation (**13**) does not make $N=0$ at $z=h$, and so does not represent the exact solution. If a suitable solution of the homogeneous equation corresponding to (**5**) is added then this boundary condition can be satisfied. The resulting expression is rather complicated and can be evaluated only approximately. The physical basis of the approximation may be represented as follows : suppose a uniform stream to approach a plane electrode at right angles to the stream, the concentration at a large distance from the electrode being N_0.

At a distance s from the electrode, the concentation is

$$N=N_0\{1-\exp(-2\lambda_1 s)\}$$

and the flux of ions to an area $\mathrm{d}S$ of the electrode is

$$\left(D_1\frac{\mathrm{d}N}{\mathrm{d}S}\right)_{s=0}\mathrm{d}S=2\lambda_1 D_1 N_0\mathrm{d}S=W_1N_0\mathrm{d}S,$$

which is the flux across an element $\mathrm{d}S$ of a geometrical plane normal to the undisturbed stream.

In an ionic stream $2\lambda_1=W_1/D_1\simeq 40E$, where E is the electron field strength in volt cm^{-1} and W_1 and D_1 are expressed respectively in cm sec^{-1} and cm^2 sec^{-1}.

If $E=1$ V cm^{-1}, it follows that $N\simeq\frac{2}{3}N_0$ at a distance of 0·5 mm from the electrode.

It follows that the distribution in the stream of ions of equation (**13**) in the vicinity of the electrode at $z=h$ is given closely by $N\{1-\exp-2\lambda_1(2h-z)\}$ and that the flux to the surface element $\mathrm{d}S$ of the electrode at a point (ρ,h) is $NW_1\mathrm{d}S$, where N is given by equation (**13**) with $z=h$. This, as explained above, is the procedure that was adopted in practice.

The exact mathematical treatment is most conveniently expressed in terms of Fourier Bessel transforms. The approximate solution (**13**) can be written in the alternative form

$$N=\frac{\alpha_a WS}{8\pi DD_1}e^{\lambda z}\int_0^\infty k\mathrm{d}k\frac{J_0(k\rho)}{\sqrt{(\mu^2+k^2)}}\left\{\frac{\exp(z-2h)\sqrt{(\mu^2+k^2)}}{\alpha\lambda+(\lambda-\lambda_1)\{\lambda+\sqrt{(\mu^2+k^2)}\}}\right.$$
$$\left.-\frac{\exp\{-z\sqrt{(\mu^2+k^2)}\}}{\alpha\lambda+(\lambda-\lambda_1)\{\lambda-\sqrt{(\mu^2+k^2)}\}}\right\}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{13'})$$

The expression (**13′**) is defined for $z\geqslant 0$. For $z<0$, the second term inside the bracket must be replaced by $\exp\{z\sqrt{(\mu^2+k^2)}\}/[\alpha\lambda+(\lambda-\lambda_1)\{\lambda+\sqrt{(\mu^2+k^2)}\}]$. The two parts of the solution so obtained do not join up smoothly at $z=0$. But suitable choice of the complementary function for $z<0$ will ensure that the solution is continuous and satisfies the correct boundary condition at $z=-\infty$. So strictly speaking (**13′**) is not equivalent to (**13**) but rather to

$$\bar{N}=-\frac{S\alpha_a\lambda}{4\pi D_1(\lambda_1-\lambda)}\exp\left(\frac{\alpha\lambda z}{\lambda_1-\lambda}\right)\left\{\int_z^\infty \mathrm{d}z'\exp\left(1-\frac{\alpha}{\lambda_1-\lambda}\right)\lambda z'\,\frac{\exp(-\mu r')}{r'}\right.$$
$$\left.+\int_{-\infty}^{z}\mathrm{d}z'\exp\left(1-\frac{\alpha}{\lambda_1-\lambda}\right)\lambda z_1\,\frac{\exp(-\mu r'')}{r''}\right\}$$
$$=N^{+}+N^{-}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{13''})$$

It can be verified that the following choice of complementary function is suitable :

$$N_c=-\frac{\alpha_a WS}{8\pi DD_1}\int_0^\infty k\mathrm{d}k\frac{J_0(k\rho)}{\sqrt{(\mu^2+k^2)}}$$
$$\times\exp\left[(z-h)\{\lambda_1+\sqrt{(\lambda_1^2+k^2)}\}+h\{\lambda-\sqrt{(\mu^2+k^2)}\}\right]$$
$$\times\left[\frac{1}{\alpha\lambda+(\lambda-\lambda_1)\{\lambda+\sqrt{(\mu^2+k^2)}\}}-\frac{1}{\alpha\lambda+(\lambda-\lambda_1)\{\lambda-\sqrt{(\mu^2+k^2)}\}}\right]. \qquad \ldots\ldots\ldots\ldots \quad (\mathbf{14})$$

The ion current at the anode is now

$$-D_1\frac{\partial}{\partial z}(N+N_c)\Big|_{z=h}.$$

From (**13**),

$$-D_1\frac{\partial N}{\partial z}\Big|_{z=h}=-\frac{\lambda\alpha D_1}{\lambda_1-\lambda}N\Big|_{z=h}.$$

The calculation of $\partial N_c/\partial z\,|_{z=h}$ can be carried out approximately if $\lambda_1\gg\mu$ or $\lambda_1\simeq\mu$. The integrand contains a factor $\lambda_1+\sqrt{(\lambda_1^2+k^2)}$ on differentiating with respect to z, and then putting $z=h$. If $\lambda_1\gg\mu$, we write $\lambda_1+\sqrt{(\lambda_1^2+k^2)}=\lambda_1+\sqrt{\{(\lambda_1^2-\mu^2)+(\mu^2+k^2)\}}$, and expand in powers of $(\mu^2+k^2)/(\lambda_1^2-\mu^2)$. To the first order we have

$$\lambda_1+\sqrt{(\lambda_1^2+k^2)}=\lambda_1+\sqrt{(\lambda_1^2-\mu^2)}+\frac{\mu^2+k^2}{\sqrt{(\lambda_1^2-\mu^2)}}-\frac{(\mu^2+k^2)\sqrt{(\lambda_1^2+k^2)}}{\sqrt{(\lambda_1^2-\mu^2)}\{\sqrt{(\lambda_1^2+k^2)}+\sqrt{(\lambda_1^2-\mu^2)}\}}.$$

Neglecting the last term, this gives the approximate expression

$$-D_1\frac{\partial N_c}{\partial z}\Big|_{z=h}=D_1\{\lambda_1+\sqrt{(\lambda_1^2-\mu^2)}\}N\Big|_{z=h}+\frac{D_1}{\sqrt{(\lambda_1^2-\mu^2)}}\,\frac{\partial^2}{\partial z^2}\mathrm{e}^{-\lambda z}N\Big|_{z=h}$$

$$\simeq 2\lambda_1 D_1 N\Big|_{z=h}+\frac{D_1}{\lambda_1}\left(\frac{\partial^2}{\partial z^2}\mathrm{e}^{-\lambda z}N\Big|_{z=h}-\tfrac{1}{2}\mu^2 N\Big|_{z=h}\right), \quad \ldots \quad \mathbf{(15)}$$

agreeing with the physical argument given earlier.

If $\lambda_1\simeq\mu$, we write

$$\lambda_1+\sqrt{(\lambda_1^2+k^2)}=\lambda_1+\sqrt{(\mu^2+k^2)}+\frac{\lambda_1^2-\mu^2}{2\sqrt{(\mu^2+k^2)}}$$

$$+\frac{(\lambda_1^2-\mu^2)^2}{2\sqrt{(\mu^2+k^2)}}\;\frac{1}{\{\sqrt{(\mu^2+k^2)}+\sqrt{(\lambda_1^2+k^2)}\}^2}.$$

This is an expansion in powers of $(\lambda_1^2-\mu^2)/(\mu^2+k^2)$, and, to the first order, we can neglect the last term. Then, with the notation introduced above,

$$-D_1\left(\frac{\partial N_c}{\partial z}\right)_{z=h}=\lambda_1 D_1 N\Big|_{z=h}-D_1\frac{\partial}{\partial z}\mathrm{e}^{-\lambda z}(N^+ + N^-)\Big|_{z=h}$$

$$+\tfrac{1}{2}D_1(\lambda_1^2-\mu^2)\left(\int_h^\infty \mathrm{d}z'N^+ - \int_{-\infty}^h \mathrm{d}z'N^-\right). \quad \ldots\ldots\ldots\ldots \quad \mathbf{(15')}$$

The expressions **(15)**, **(15′)** can be calculated numerically without too much difficulty.

IV. REFERENCES

HUXLEY, L. G. H. (1959).—*Aust. J. Phys.* **12**: 171.

HUXLEY, L. G. H., CROMPTON, R. W., and BAGOT, C. H. (1959).—*Aust. J. Phys.* **12**: 303.

THE DESIGN OF PHOTOGRAPHIC OBJECTIVES OF THE TRIPLET FAMILY

II. THE INITIAL DESIGN OF COMPOUND TRIPLET SYSTEMS

By F. D. Cruickshank*

[*Manuscript received September* 28, 1959]

Summary

Part II. The general method developed for the type 111 triplet is extended to the initial design of triplet systems having compound components. The possibility of replacing a single lens by an equivalent doublet or triplet allows the single lens to be regarded as not being restricted to values of N and V associated with known glass types. It follows that the glass parameters α, β, γ, and ξ may be treated as continuously variable over limited ranges. The general effects of the variation of these parameters on the powers and separations of the basic initial solutions are shown in a set of diagrams. As an example, the initial design of a triplet with compound members is given in detail.

I. Introduction

In Part I of this series (Cruickshank 1958) the basic equations and procedures have been given for the initial design of the parent triplet objective, type 111, for a given selection of glasses and prescribed values of the power, the paraxial chromatic aberrations, and the third-order monochromatic aberrations. The purpose of the present paper is to extend the general method to the initial design of other members of the triplet family (Fig. 1 of Part I) having compound components. The first step necessary is to consider the equivalent glass properties of such components.

II. Equivalent Doublets and Triplets

Consider the thin lens (φ,N,V) of power φ made from glass of refractive index N and V-number V. Suppose this lens is to be replaced by n thin lenses $(k_1\varphi,N_1,V_1)$, $(k_2\varphi,N_2,V_2)$, . . . , $(k_n\varphi,N_n,V_n)$ in contact, the replacement being made in such a way that the compound group has the same power, the same Petzval sum, and the same paraxial chromatic aberrations. This requires that

$$\sum_{j=1}^{n} k_j = 1, \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

$$\sum_{j=1}^{n} k_j\omega_j = \omega, \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

$$\sum_{j=1}^{n} k_j\psi_j = \psi, \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

where $\omega=1/N$, $\omega_j=1/N_j$, $\psi=1/V$, $\psi_j=1/V_j$. Two particular cases arise for detailed consideration.

---

* Department of Physics, University of Tasmania, Hobart.

The first of these is the case in which the replacing group is a doublet, i.e. the thin lens (φ, N, V) is replaced by the thin lenses $(k_1\varphi, N_1, V_1)$, $(k_2\varphi, N_2, V_2)$ in contact. The parameters k_1, k_2 must be chosen so that

$$k_1+k_2 \quad =1, \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{4})$$

and either

$$\omega_1 k_1+\omega_2 k_2=\omega, \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{5})$$

or

$$\psi_1 k_1+\psi_2 k_2=\psi. \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{6})$$

Many possibilities now arise. Suppose, for example, that the single *positive* thin lens (φ, N, V) is replaced by a doublet consisting of a thin *positive* lens $(k_1\varphi, N_1, V_1)$ and a thin *negative* lens $(k_2\varphi, N_2, V_2)$ in contact. Then k_2 will be negative, $|k_2/k_1|<1$, and both N_1 and N_2 must be greater or less than N according as $N_2 \gtrless N_1$. Similarly, both V_1 and V_2 must be greater or less than V according as $V_2 \gtrless V_1$.

Looking at the matter from the opposite point of view, suppose a thin doublet of positive power is constructed from the positive thin lens $(k_1\varphi, N_1, V_1)$ and the negative thin lens $(k_2\varphi, N_2, V_2)$ in contact. This doublet is equivalent as regards power, Petzval sum, and paraxial chromatic aberrations to a single thin lens (φ, N, V) where N and V are given through equations (**5**), (**6**). These values N, V will not correspond in general to those of any known glass and may be varied widely by intelligent choice of the two glasses and the value of k_2/k_1. In particular, by selecting a pair of glasses of the same index for some mean wavelength and varying k_2/k_1, the effective value of V may be varied over a wide range without change of index. Such an iso-N doublet can be very useful. For example, if we combine a positive thin lens of the DBC glass 658508 with a negative thin lens of EDF 654335 then the equivalent thin lens has an effective N-value of 1·66 approximately and an effective V-value anywhere between 50 and infinity depending on the value of k_2/k_1. Similarly, by choosing an iso-V pair of glasses the effective value of N may be varied substantially without change of V. More generally, the choice of glass pairs in which both N and V are dissimilar allows the effective values of N and V to be changed simultaneously over a considerable range.

Consider next the special case in which the replacing group is a triplet. Thus suppose the thin lens (φ, N, V) is replaced by a compound triplet consisting of the three thin lenses $(k_1\varphi, N_1, V_1)$, $(k_2\varphi, N_2, V_2)$, $(k_3\varphi, N_3, V_3)$ in contact. The replacement may now be made in such a way that the triplet has the same power, the same Petzval sum, and the same chromatic aberrations as the singlet it replaces. Hence the equations (**1**)–(**3**) become

$$k_1+k_2+k_3 \quad =1, \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{7})$$

$$\omega_1 k_1+\omega_2 k_2+\omega_3 k_3=\omega, \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{8})$$

$$\psi_1 k_1+\psi_2 k_2+\psi_3 k_3=\psi, \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{9})$$

whence

$$k_1=\Delta_1/\Delta\,; \quad k_2=\Delta_2/\Delta\,; \quad k_3=\Delta_3/\Delta, \qquad \ldots\ldots\ldots \quad (\mathbf{10})$$

where

$$\Delta_1=\begin{vmatrix} 1 & 1 & 1 \\ \omega & \omega_2 & \omega_3 \\ \psi & \psi_2 & \psi_3 \end{vmatrix}, \qquad \Delta_2=\begin{vmatrix} 1 & 1 & 1 \\ \omega_1 & \omega & \omega_3 \\ \psi_1 & \psi & \psi_3 \end{vmatrix},$$

$$\Delta_3=\begin{vmatrix} 1 & 1 & 1 \\ \omega_1 & \omega_2 & \omega \\ \psi_1 & \psi_2 & \psi \end{vmatrix}, \qquad \Delta=\begin{vmatrix} 1 & 1 & 1 \\ \omega_1 & \omega_2 & \omega_3 \\ \psi_1 & \psi_2 & \psi_3 \end{vmatrix}.$$

It is thus formally possible to design a triplet component which is equivalent as regards power, Petzval sum, and chromatic aberrations to a thin lens having, within a certain range, any value of N, and, independently thereof, any value of V. Therefore in the development of a photographic objective we may ignore the restrictions imposed by the existence of only a limited number of glass types if we are prepared to use cemented doublets and triplets to replace components which are single thin lenses in the initial stages of the design. For example, if for some reason one requires for the negative component of a triplet a glass for which $N=2\cdot 0$ and $V=15$, it can be assumed that this fictitious glass exists and in the next stage of design these effective values can be achieved by replacing the fictitious component by an actual cemented triplet.

In the design of triplet objectives of type 111 we are limited to available glass types for the three components and hence the values of the glass parameters,

$$\alpha=V_a/V_b\,, \qquad \xi=V_a/V_c\,,$$
$$\beta=N_a/N_b\,, \qquad \gamma=N_a/N_c\,,$$

are not continuously variable quantities. In the light of the considerations of this section, however, those members of the triplet family of objectives which have compound components may be considered at first as simple type 111 triplets in which one or more components have fictitious glass constants. This simplifies the problem of designing such systems as the theory of the initial design of type 111 triplets given in Part I is immediately applicable since each equivalent simple triplet has the same power, the same corrector power χ, and the same Petzval sum and chromatic aberrations as the compound system to which it is equivalent. The appearance in this way of glass constants which do not correspond necessarily to known types removes substantially the restrictions on the glass parameters and it is neither artificial nor without physical meaning to regard α, β, ξ, and γ as continuously variable over limited ranges at least.

III. THE EFFECTS OF THE VARIATION OF THE GLASS PARAMETERS

The first stage in the development of the design of a triplet objective is the determination of the powers φ_a, φ_b, φ_c and the separations t_1, t_2 of the three thin lenses a, b, c constituting the initial arrangement of the system. From the point of view of this first stage the variety of constructions in the triplet family is largely the consequence of the variation of the glass parameters α, β, ξ, and γ. If an understanding of the range of properties of the whole family of objectives is to be reached, or, indeed, an understanding of the detailed properties of any one member, it is necessary to know the effect on the initial solution of the variation of α, β, ζ, and γ. It is not assumed that this is the whole story as regards glass selection, because each of these parameters is a ratio and therefore any given values of the parameters might be achieved in a variety of ways. However, it represents a systematic approach to what is a complicated problem, and, once this general approach is outlined, it may be supplemented by a closer study of the effects of achieving any of these parameters in different ways in connexion with different members of the triplet family.

When the changes in the powers and separations due to given changes in the glass parameters are known there remains still the problem of determining whether the new basic solution is more favourable or less favourable than its predecessor. This is effectively the second stage of the design. Broadly speaking, an initial solution must be scrutinized in three different ways. The inherent characteristics of the system in respect of spherical aberration of all orders must be examined, then the asymmetry errors, i.e. the comas of all types and orders thereof, and finally those characteristics which control the field correction, i.e. the astigmatism of different orders and what is generally of over-riding importance in photographic objectives the oblique spherical aberrations.

In this laboratory extensive use is made of the aberration coefficients introduced by Buchdahl (1954). The general method developed for the use of these coefficients is described in another paper by Cruickshank and Hills (1960). Anyone who has used these coefficients will readily concede that they provide the most powerful means available at present in geometrical optics for predicting the image quality of a system and for analysing the way in which the particular construction achieves its performance. The laboratory's computing room furnishes a standard service of the computation of all the aberration coefficients of third, fifth, and seventh orders for any system. Formerly these coefficients were computed on electric desk calculating machines, but recently Ford (1959) has arranged the computation on a digital computer. In these computations the total value of each aberration coefficient of the complete system is obtained as the sum of a set of partial coefficients, one for each surface of the system. If, for example, a variation of the basic parameters leads to a general reduction of the partial coefficients of the spherical aberration at the several surfaces of the system, it is permissible to infer that this represents a step towards the development of reduced zonal spherical aberration and hence the possibility of increased aperture for the system if needed. One has only to compare the partial aberration coefficients of a few first class objectives designed for apertures of, say $f/3 \cdot 5$, $f/2 \cdot 5$, $f/2$, and $f/1 \cdot 5$ respectively, to see that this is so. No successful high aperture systems are developed from initial arrangements with characteristically large partial coefficients of spherical aberration. Now, broadly speaking, reduced partial coefficients of spherical aberration go hand in hand with reduced powers of the components, which in turn lead to shallower curves throughout the system. Of course, reduced powers cannot be sought without consideration of the separations of the components which may become so small that the system cannot be built or so large that there is an unacceptable degree of vignetting. The criteria by which the inherent properties of a system in respect of asymmetry errors and field errors are to be judged are similarly based on an examination of the appropriate aberration coefficients. The calculation of the aberration coefficients provides an analysis of how these aberrations arise within any particular system, but just how these are linked with simpler quantities in the initial solution requires deeper investigation. Further work is in progress in this connexion, in the course of which it is planned to make a broad study of the characteristic distribution of these aberration coefficients within the various types of construction which constitute the triplet family.

(*a*) *Objectives for which* $\xi=\gamma=1$

Consider first those objectives in which the effective glass constants of the front and back components are the same. It is common practice to construct type 111 objectives from two glasses, using the same crown glass for both positive lenses. This makes $\xi=1=\gamma$ and leaves only α and β as the effective glass parameters. Extending this idea to triplets with compound components it is possible to make systems like types 212, 222, 313, 323, etc. on this pattern provided the effective N and V values for the front and back components are the same.

Choosing some typical values for the aberration residuals which enter into the fundamental equations

$$\chi=-0\cdot 25, \qquad L=+0\cdot 20,$$
$$P=+0\cdot 50, \qquad T=+0\cdot 05,$$

the variation of the powers and separations of the thin lenses in the initial arrangement have been calculated for a range of values of α and β that are quite practical. Figures 1 (*a*)–1 (*f*) present a summary of the results.

It is immediately clear that increasing the value of β leads to substantial reductions in the powers of the first two lenses and to a lesser degree reduces the power of the back lens also. Intelligent choice of β can lead, therefore, to shallower curves throughout the system, with consequent reductions in the values of the partial aberration coefficients at the individual surfaces of the system. As β increases from lower values the front airspace at first widens but subsequently diminishes, while the rear airspace decreases steadily throughout the range. For any value of β the $|\varphi_b|$ versus α curve exhibits a shallow minimum somewhere towards the higher end of the α range. In addition, φ_a is decreasing steadily as α increases, while the accompanying increase in φ_c is not nearly so large. This may suggest, in general, the selection of an α-value near to that at which $|\varphi_b|$ attains its minimum. However, the very considerable increases in t_1 and t_2 as the value of α increases must be weighed against this.

There is another important matter to be considered in connexion with the choice of α, and that is the correction of the spherical aberration of the system. It was shown in Part I of this study that, once the glass types and values of desired residuals have been chosen for a triplet, the shapes of the three components can be adjusted to secure control of the primary astigmatism, distortion, and coma, while the primary spherical aberration, σ_1, is controlled by the parameter χ, the power of the corrector of the system. The whole placing of the σ_1 versus χ parabola is often controlled by the choice of α as is seen in Figure 2.

It cannot be assumed, however, that everything is necessarily satisfactory provided the σ_1 versus χ parabola comes near to or intersects the χ-axis. As α decreases and the corresponding parabola drops towards the χ-axis, the general magnitudes of the partial aberration coefficients of the surfaces increase. Varia-

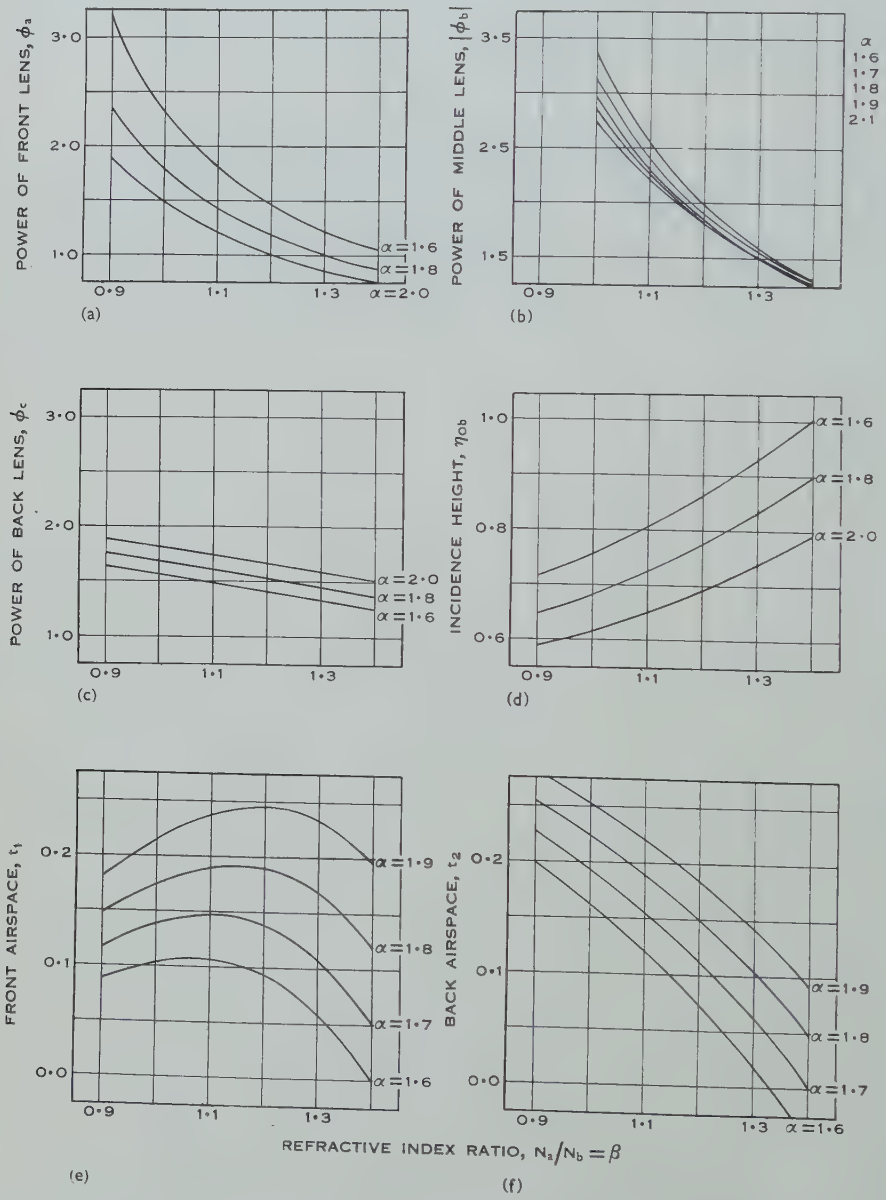


Fig. 1.—The effects of the variation of β on the powers and separations of the thin components of unit power triplets having different values of α are shown. In all triplets $\xi=1$, $\gamma=1$, and $\chi=-0\cdot25$. The residuals of Petzval curvature and the chromatic aberrations are in all cases $P=0\cdot50$, $L=0\cdot20$, $T=0\cdot05$.

tion of χ for a fixed α-value leaves the partial coefficients of spherical aberration of the negative lens almost unchanged, while the partial coefficients of the positive lenses change considerably in accordance with the parabolic law. On the other hand, the partial coefficients for the negative lens, and hence also those for the positive lenses, increase very substantially as α is decreased. This is due principally to the effect of α upon the front airspace. A decrease in α lowers the front airspace, raises η_{0b}, and hence the spherical aberration coefficients at the surface of the negative lens. If, then, it is desired to secure as high an aperture as is consistent with a certain type of triplet construction, it is important to consider the location of the σ_1 versus χ parabola during selection of glass for the system, for the lower order partial aberration coefficients largely determine the magnitude of the higher order aberration coefficients of the system. A thorough study of Figures 1 and 2 will enable the reader to grasp some of the essential significance of glass selection in the design of this class of triplets.

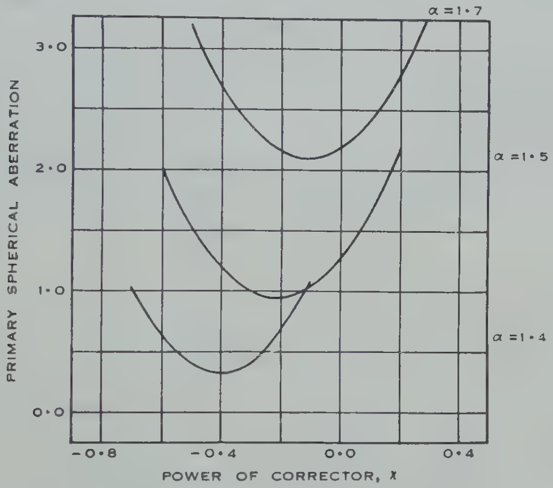


Fig. 2.—This figure shows the effect of the variation of α on the placing of the σ_1 *v.* χ parabola in a type 212 triplet.

(*b*) *Objectives for which* $\xi \neq \gamma$. *The General Case*

Apart from the construction of a series of three-dimensional models, it is difficult to present the effects of the simultaneous variation of the four glass parameters on the powers and separations in the general case of three-glass objectives. What has been done, therefore, is firstly to set $\beta = 1 = \gamma$ and investigate the effect of the variation of ξ. The results are presented for selected values of α in Figures 3 (*a*)–3 (*f*). It is seen that an increase in the value of ξ increases the powers of the lenses in the corrector system and slightly decreases the power of the back component, the changes being almost linear over this range in each case. Over against this are fairly rapid decreases in both airspaces. Secondly, setting $\beta = 1 = \xi$, the variation of γ has been investigated over the range 0·8–1·2. The results are presented in Figures 4 (*a*)–4 (*f*), from which it is seen that an increase of γ increases all powers, decreases the front airspace, and increases the back airspace. At smaller values of α the effect on the front airspace is very small. It is noticeable in all these variations that the effects

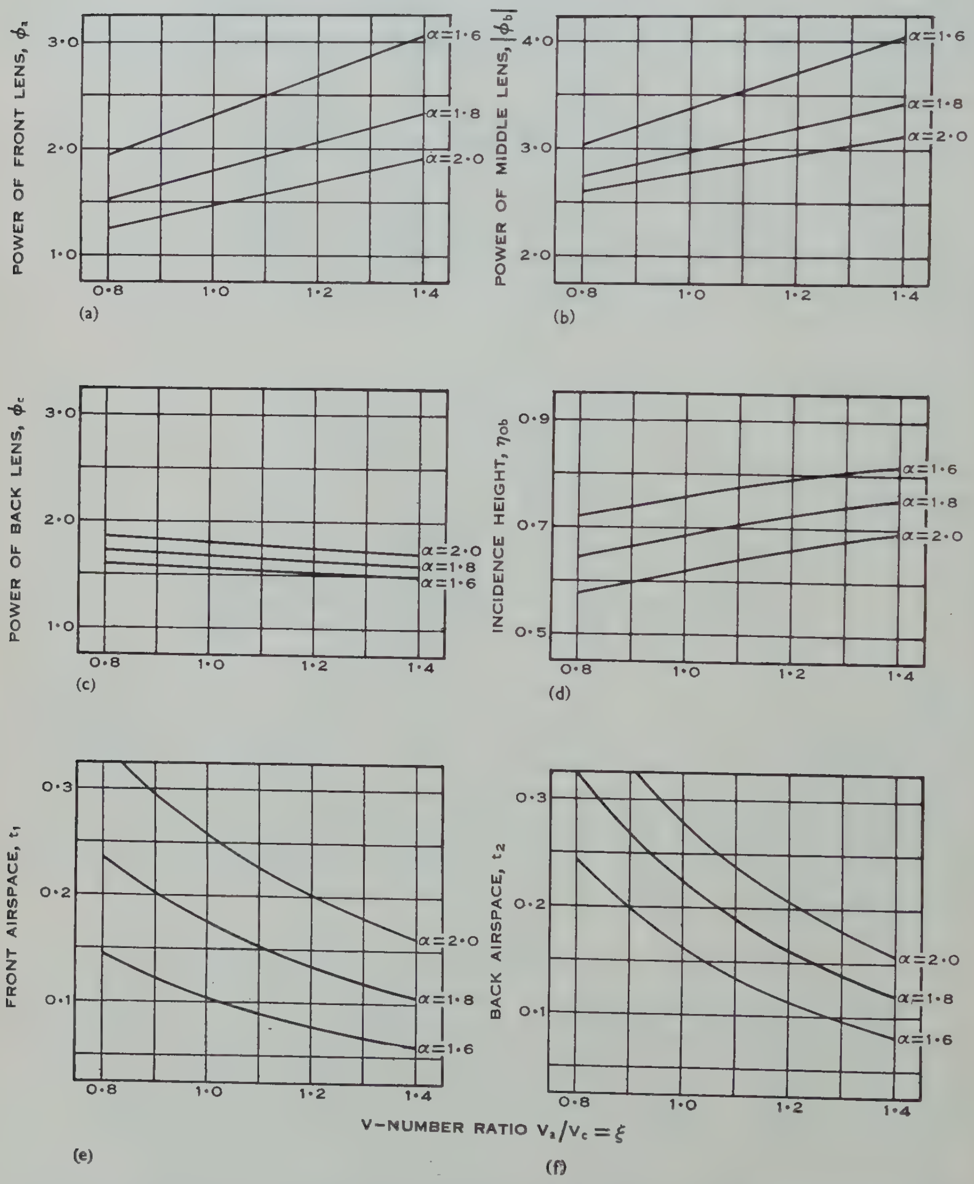


Fig. 3.—These figures show the effects of the variation of ξ on the powers and separations of the thin components of unit power triplets having different values of α. In each triplet $\beta=1$, $\gamma=1$, and $\chi=-0\cdot25$. The residuals of Petzval curvature and chromatic aberration are $P=0\cdot50$, $L=0\cdot20$, $T=0\cdot05$.

on the powers of the first two lenses are very much greater than on the power of the back component. This component behaves as a fairly stable feature of the construction, while the front portion of the system may undergo wide changes. This strengthens the view adopted by the writer that the first two lenses fulfil the function of a corrector system.

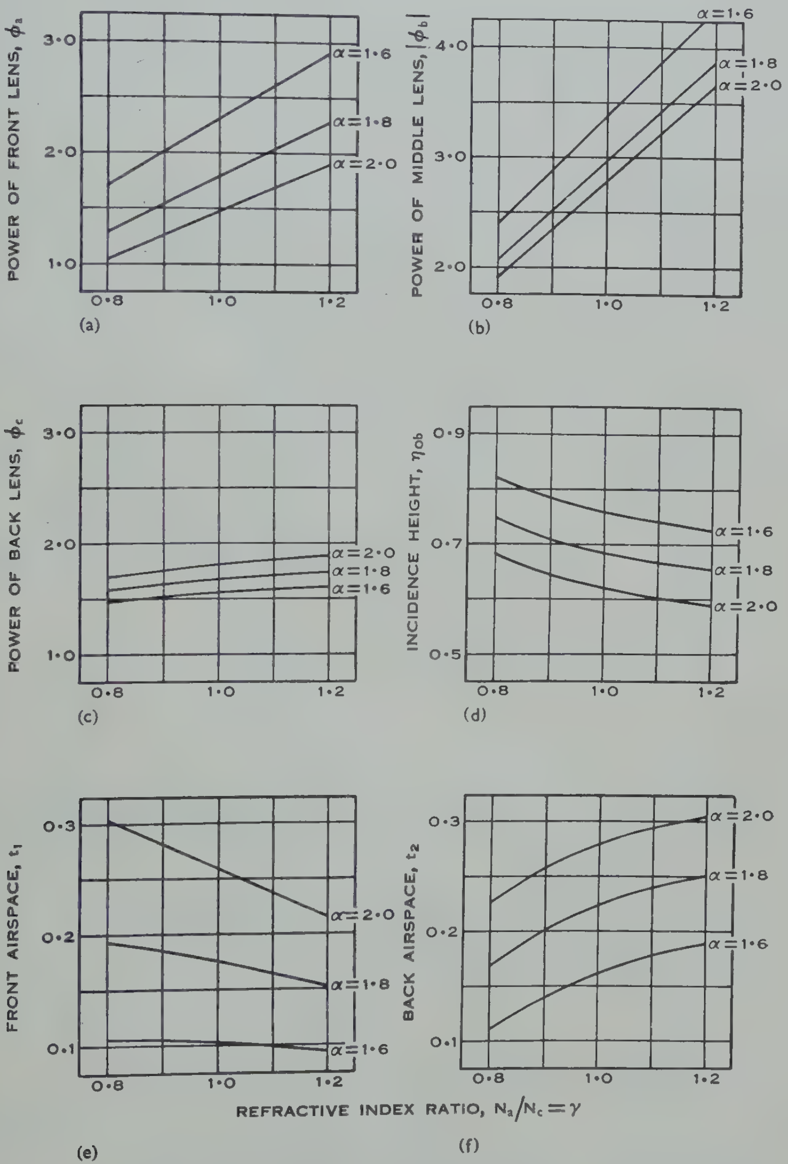


Fig. 4.—These figures show the effect of the variation of γ on the powers and separations of the thin components of unit power triplets having different values of α. In each triplet $\beta=1$, $\xi=1$, and $\chi=-0\cdot25$. The residuals selected are $P=0\cdot50$, $L=0\cdot20$, and $T=0\cdot05$.

IV. AN EXAMPLE OF INITIAL TRIPLET DESIGN

It will be useful now to discuss in detail an actual example of an initial design to illustrate the general method. Suppose we consider what could be achieved in a type 212 lens in which the front and rear components are iso-index doublets. We choose from the glass catalogue the approximately iso-index pair of glasses 1·65695, 50·80 and 1·6538, 33·55 for this purpose. The equivalent refractive index of a doublet from this pair will be approximately 1·66 but the equivalent V-number will depend on the relative values of k_1 and k_2. Equations (**4**), (**6**) show, for instance, that for the range $0 \leqslant |k_2/k_1| \leqslant 0 \cdot 6$ the corresponding equivalent V-number range is $50 \cdot 8 \leqslant V \leqslant 222$, so that we may give V almost any value we please.

Figures 1 (*a*)–1 (*c*) show that the powers of all components are reduced by making β as high as possible. In addition, the powers of the first two components are reduced by using a high value of α. Having fixed the value of N_a at approximately 1·66 by the choice of the iso-index pair, there remains only one way of increasing β and that is the choice of a low index glass for the central negative lens. Accordingly we choose the BSC glass 1·5095, 64·4 for this. It may seem rather unusual to choose such a crown glass for the main dispersive member of the triplet, but at least it does provide the increased value of β which seems desirable. We do not know yet what value of α will be required to make the correction of spherical aberration possible. Initially, a high value of α is desirable to give further reduction in the powers of the first two components. Quite arbitrarily we choose $\alpha=2$ and see what happens. Now $\alpha=V_a/V_b$ whence $V_a=2\times 64 \cdot 4=128 \cdot 8$. Equations (**4**), (**6**) therefore give $k_2/k_1=-0 \cdot 54083$ and $k_1=2 \cdot 17785$, $k_2=-1 \cdot 17785$. With these k values equation (**5**) gives $N_a=1 \cdot 66068$ and hence $\beta=N_a/N_b=1 \cdot 1002$.

Since we are using the same iso-index pair for the back doublet, the equivalent refractive index of this doublet will be also approximately 1·66 and hence $\gamma=N_a/N_c$ will be very close to unity. It only remains, then, to settle the value of ξ. Figures 1 (*e*) and 1 (*f*), which are drawn for $\xi=1$, $\gamma=1$, show that the choice of $\alpha=2$ and $\beta=1 \cdot 100$ results in high values for the airspaces if $\xi=1$. Figures 3 (*e*) and 3 (*f*) show, however, that if ξ is increased the airspaces will be reduced substantially. Figures 3 (*a*)–3 (*c*) show also that this reduction in the airspaces will be obtained at the expense of some increase in the powers of the first and second components. A compromise is therefore called for and so, tentatively, we put $\xi=1 \cdot 5$. Since $\xi=V_a/V_c$, this gives $V_c=128 \cdot 8/1 \cdot 5=85 \cdot 867$ and thus $k_1=1 \cdot 79439$, $k_2=-0 \cdot 79439$, and $k_2/k_1=-0 \cdot 44270$. It then follows that the equivalent index of the back doublet is $N_c=1 \cdot 65946$ and thus $\gamma=N_a/N_c=1 \cdot 0007$.

The next matter is to select the values of P, L, and T. Putting $P=0 \cdot 6$ makes the Petzval sum $P/N_a=0 \cdot 6/1 \cdot 66068=0 \cdot 3613$. This is fairly high but can be tolerated if the field angle required is not too great. If we make L slightly positive the longitudinal chromatic aberration will be corrected at an outer zone of the aperture rather than paraxially. We therefore put $L=0 \cdot 50$, and similarly $T=0 \cdot 045$. Now, using equations (**21**)–(**32**) of Part I, initial solutions are made for three values of the corrector power, say $\chi=-0 \cdot 25$, $-0 \cdot 375$, and $-0 \cdot 50$.

The resulting equivalent thin solutions are summarized in the first five lines of Table 1. Using the values $k_2/k_1=-0\cdot54083$ and $-0\cdot44270$ for the front and back doublets respectively, the powers of the members of these doublets are calculated in each case and are shown in the next four lines of Table 1. This completes the calculation of the powers and separations of the thin components for these three systems.

TABLE 1

DETAILS OF DESIGN EXAMPLE DURING THIN LENS STAGE

χ	$-0\cdot250$	$-0\cdot375$	$-0\cdot500$
φ_a	$1\cdot2723$	$1\cdot2016$	$1\cdot1322$
φ_b	$-2\cdot1306$	$-2\cdot1710$	$-2\cdot2156$
φ_c	$1\cdot6705$	$1\cdot7857$	$1\cdot9040$
t_1	$0\cdot2244$	$0\cdot2279$	$0\cdot2326$
t_2	$0\cdot1351$	$0\cdot1169$	$0\cdot1023$
φ_{a1}	$2\cdot7708$	$2\cdot6168$	$2\cdot4657$
φ_{a2}	$-1\cdot4985$	$-1\cdot4152$	$-1\cdot3335$
φ_{c1}	$-1\cdot3270$	$-1\cdot4184$	$-1\cdot5124$
φ_{c2}	$2\cdot9975$	$3\cdot2041$	$3\cdot4164$
S_{a1}	$0\cdot1853$	$0\cdot2276$	$0\cdot2718$
S_{a2}	$0\cdot4992$	$0\cdot4213$	$0\cdot3401$
S_b	$-0\cdot2744$	$-0\cdot2027$	$-0\cdot1232$
S_{c1}	$-1\cdot9954$	$-1\cdot9220$	$-1\cdot8501$
S_{c2}	$0\cdot3324$	$0\cdot2998$	$0\cdot2678$
σ_1	$0\cdot9350$	$0\cdot8910$	$0\cdot9249$
σ_2	$+0\cdot0007$	$-0\cdot0002$	$-0\cdot0001$
σ_3	$-0\cdot0005$	$-0\cdot0005$	$-0\cdot0004$
$2\sigma_4$	$0\cdot3613$	$0\cdot3613$	$0\cdot3613$
σ_5	$0\cdot0000$	$0\cdot0000$	$0\cdot0000$

We come now to the determination of the shapes of the thin lenses. Assuming that the two doublets will be cemented, the shapes of the lenses 1, 2 of each doublet must satisfy the relation

$$\varphi_1(S_1-1)/(N_1-1)=\varphi_2(S_2+1)/(N_2-1). \quad \ldots\ldots\ldots \quad (\mathbf{11})$$

There are thus only three independent shapes, say, S_{a1}, S_b, and S_{c2}. The condition for zero third-order astigmatism is now set up using equation (**44**) of Part I, the diaphragm being assumed to be coincident with the central negative lens *b*. In the case of $\chi=-0\cdot375$, for example, we obtain

$$2\cdot8746S_{a1}^2-6\cdot1068S_{a1}+S_{c2}^2+3\cdot4401S_{c2}+0\cdot11992=0.$$

Similarly, the condition for zero third-order distortion is derived from equation (**45**) of Part I, and for $\chi=-0\cdot375$ we find

$$-4\cdot3148S_{a1}^2+14\cdot1855S_{a1}+S_{c2}^2+5\cdot8447S_{c2}-4\cdot8468=0,$$

the solution of these two quadratics giving

$$S_{a1}=0\cdot22759, \qquad S_{c2}=0\cdot29975.$$

Equation (**11**) then gives $S_{a2}=0\cdot42134$, and the corresponding relation for the back doublet gives $S_{c1}=-1\cdot92195$. The condition for zero third-order coma is then written down from equation (**43**) of Part I and this gives S_b directly. As a check on the calculations up to this point the third-order aberrations are calculated for the thin systems thus obtained. The computations up to this stage are summarized in the third and fourth blocks of Table 1. Assuming a quadratic relation between the third-order spherical aberration σ_1 and the corrector power χ, the three values of σ_1 in Table 1 give

$$2\cdot4896\chi^2+1\cdot9074\chi+1\cdot2562=\sigma_1,$$

which should be roughly plotted as a guide to the possible correction of spherical aberration.

The next stage involves the replacement of the thin lenses by lenses of finite thickness. To do this we introduce axial thicknesses sufficient to permit a reasonable aperture, say, $f/2\cdot3$, and hence put $t_{a1}=0\cdot130$, $t_{c2}=0\cdot140$, and for the negative lenses $t_{a2}=t_b=t_{c1}=0\cdot03$. The replacements are made in the standard way which leaves the refraction of the axial paraxial ray the same as in the thin system. The third-order aberrations are now recalculated and attain the values shown in the first block of Table 2. It is apparent that substantial changes have occurred, necessitating readjustment of the shapes.

The shapes are readjusted as shown in the second block of Table 2 and the resulting aberration coefficients are given in the third block of the table. It will be seen that now the shapes have been chosen to ensure a small negative residual of primary coma and a sufficient value of σ_3 to flatten the tangential field. A small positive residual of primary distortion has been left to offset the negative secondary and tertiary distortion coefficients. The adjustment of the coefficients σ_2 to σ_5 is practically the same in each case, and the only third-order quantity that varies with χ is the spherical aberration coefficient. Assuming as usual, the quadratic relation between χ and σ_1, we find

$$1\cdot3293\chi^2+1\cdot7971\chi+0\cdot7608=\sigma_1.$$

This expression is plotted on a ten-times scale in Figure 5.

Up to this point all the computation has been done on a desk machine. Now the complete coefficients of third, fifth, and seventh orders are computed on the digital computer, the process requiring about 5 minutes computing time per system. This checks all the earlier third-order work and gives the higher order coefficients as well. In Figure 5 the variation of the secondary coefficients with χ is plotted. Each curve is based on three computed points and quadratic variation with χ is assumed. In addition the dotted curves of $10\sigma_1$ and $10\sigma_2$ are added. Because of the way the shapes are chosen the remaining primary coefficients will be practically constant, as instanced by the plot of $10\sigma_2$. They are therefore not plotted.

This diagram should be well studied. It summarizes the correction possibilities of this type of construction over a wide range of χ values. As χ approaches zero from the negative side the following properties should be noticed.

(1) *Spherical aberration.*—The third-order coefficient σ_1 is positive and increasing, while the fifth-order coefficient μ_1 is negative and decreases in absolute value : values of χ can therefore be found for which the spherical aberration can be corrected at different apertures.

(2) *Coma.*—The fifth-order circular coma specified by μ_2, μ_3 is positive and is decreasing rapidly as $\chi \rightarrow 0$; it is advantageous therefore to proceed as far in this direction as possible. The fifth-order elliptical coma (μ_7, μ_8, μ_9) is positive and, as it changes little with χ, nothing much can be done about it. Some degree of balance can be obtained by offsetting the effects of these positive coefficients by adjusting σ_2 to a negative value. This explains why σ_2 was adjusted to a small negative value earlier.

TABLE 2

DETAILS OF DESIGN EXAMPLE AFTER THE INTRODUCTION OF FINITE AXIAL THICKNESSES

χ	—0·250	—0·375	—0·500
σ_1	0·7701	0·7217	0·7423
σ_2	—0·3078	—0·2932	—0·2886
σ_3	—0·1104	—0·1204	—0·1322
$2\sigma_4$	0·3759	0·3789	0·3825
σ_5	0·0235	0·0451	0·0657
S_{a1}	0·0742	0·1384	0·2091
S_{a2}	0·7037	0·5855	0·4554
S_b	—0·3490	—0·3244	—0·3073
S_{c1}	—2·0990	—2·1363	—2·1760
S_{c2}	0·3785	0·3951	0·4128
σ_1	0·3946	0·2738	0·1945
σ_2	—0·0231	—0·0247	—0·0207
σ_3	—0·0620	—0·0604	—0·0600
$2\sigma_4$	0·3969	0·3971	0·3980
σ_5	0·0152	0·0123	0·0112
μ_1	—5·008	—7·168	—9·293
μ_2	1·548	2·850	4·489
μ_3	1·075	1·930	3·013
μ_4	—4·675	—5·595	—6·485
μ_5	—2·163	—2·576	—2·967
μ_6	—2·935	—3·316	—3·731
μ_7	2·034	1·959	1·939
μ_8	1·273	1·218	1·196
μ_9	0·7955	0·777	0·775
μ_{10}	—1·076	—1·054	—1·054
μ_{11}	—0·6291	—0·5948	—0·5678
μ_{12}	—0·0593	—0·0481	—0·0342

(3) *Curvature of field.*—The quantities which must be considered are the third-order astigmatism σ_3 and the Petzval curvature σ_4, the fifth-order

astigmatism coefficients μ_{10}, μ_{11}, and then the very important oblique spherical aberration terms involving μ_4, μ_5, μ_6. Usually it is the values of μ_4, μ_5, μ_6 which set a limit to the extent of the field of a photographic objective. In the development of this example we have already made some adjustment of the field by balancing σ_3 and σ_4 to give an approximately flat tangential field as

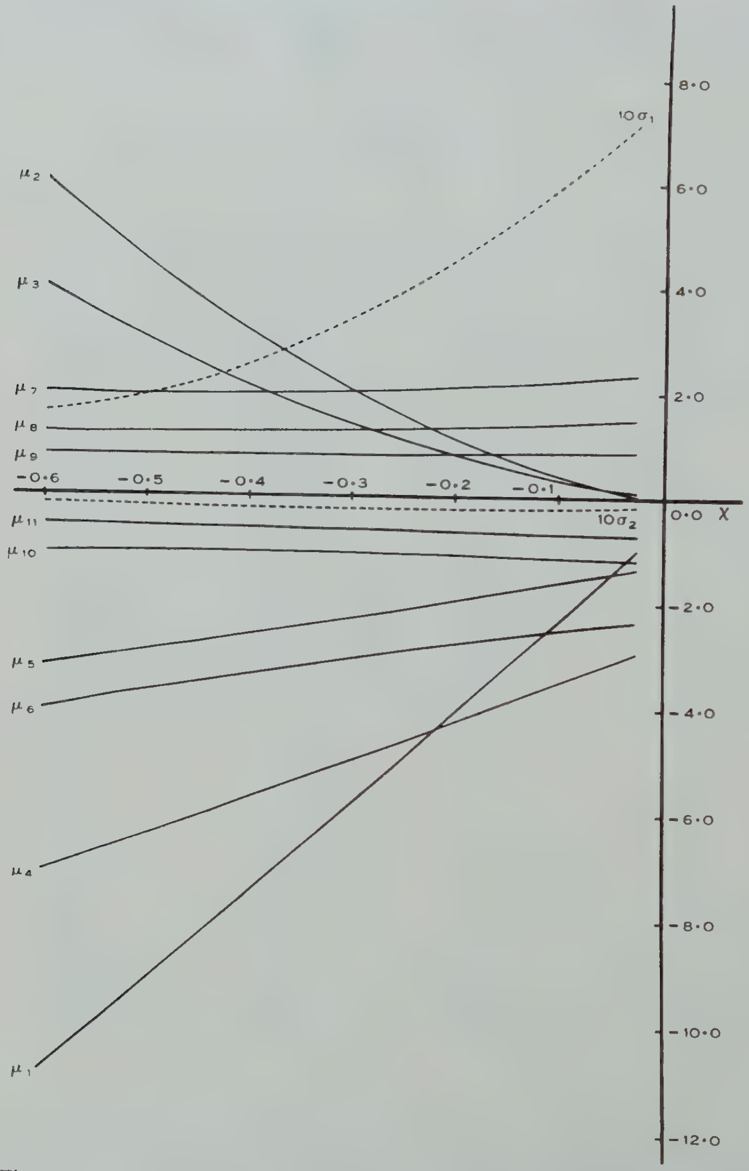


Fig. 5.—The variation of the twelve fifth-order aberration coefficients μ_1 to μ_{12}, with χ are shown for the group of type 212 triplets considered in the example. The upper dotted line shows the corresponding variation of $10\sigma_1$, the third-order spherical aberration plotted on a 10 × scale. The lower dotted line shows the variation of $10\sigma_2$, σ_2 being the third-order coma.

far as third order is concerned. Figure 5 shows that there is a significant decrease in μ_4, μ_5, μ_6 as χ tends towards zero. It will be advantageous therefore to move as far as possible in this direction for field correction as well as for the diminution of fifth-order circular coma.

Examining the spherical aberration coefficients up to the seventh order, it becomes clear that, if we select $\chi=-0\cdot27$ approximately, we can achieve spherical correction at an aperture of about $f/2\cdot3$. We can now read off the values of all the aberration coefficients of this system and thus are able to assess accurately the correction state of this system without any further detailed computation. If this is not quite what is wanted, a slight shift in χ may give a more useful system. The point to emphasize is that Figure 5 and a similar diagram showing the corresponding trends among the seventh-order coefficients summarize completely the possibilities in this type of system with this selection of glasses and residuals. If, for example, the work were repeated with some other set of glasses, a corresponding diagram would result, and a direct comparison of the two diagrams would show what advantages or disadvantages had resulted from the changes introduced. In this way areas of interest can be explored with precise knowledge of what is happening to all the aberration coefficients and why the changes are taking place.

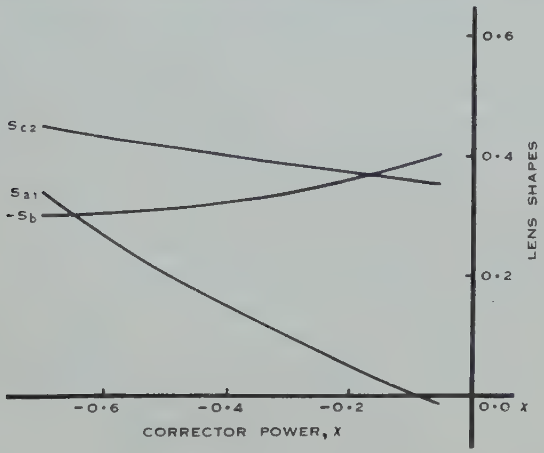


Fig. 6.—This figure shows the relations between the corrector power χ and the shapes of the thick lenses a_1, b, and c_2 which give simultaneous correction of third-order coma, astigmatism, and distortion in the type 212 triplet systems considered in the example. The shapes of the lenses for any selected χ value may be read directly from these graphs.

Suppose now that it is decided that the system with $\chi=-0\cdot270$ is the solution appropriate to a particular purpose or requirement. The detailed specification of this system can be obtained with very little further computation. The initial solution is computed with this value of χ, and this gives

$$\varphi_a=1\cdot2608,\qquad \varphi_b=-2\cdot1367,\qquad \varphi_c=1\cdot6887$$
$$t_1=0\cdot2249,\qquad t_2=0\cdot1319.$$

Using the values $k_2/k_1 = -0 \cdot 54083$ for the front doublet and $k_2/k_1 = -0 \cdot 44269$ for the back doublet, the powers of the five thin lenses of the system are obtained. Figure 6 shows the relations between the corrector power χ and the shapes of the lenses a_1, b, and c_2 for which simultaneous correction of third-order coma, astigmatism, and distortion is achieved. These curves are based on the data contained in the second block of Table 2, i.e. the final shape values used in the three computed examples to adjust the third-order coma, astigmatism, and distortion. Quadratic relations between the shapes and χ have been assumed in the curve fitting. Hence the shapes required for the lenses a_1, b, c_2 for the $\chi = -0 \cdot 270$ system may be read off directly. The values of the shapes of lenses a_2 and c_1 follow from the cementing conditions, e.g. equation (**11**). Assuming the same axial thicknesses as used already in the example and the shapes determined in the manner just described, the radii of the refracting surfaces of each lens may be computed and thus the specification completed. Running down the intersections of the curves of Figure 5 with the ordinate $\chi = -0 \cdot 270$, the aberration coefficient values expected for the system are then read off. These may then be confirmed by a 5-minute computation on the digital computer.

V. ACKNOWLEDGMENTS

The author wishes to acknowledge gratefully the help of Mrs. B. J. Brown, Senior Computer at the laboratory, in much of the computation associated with this work and in its preparation for publication. Grateful thanks are also due to Miss Jennifer James, who carried out the computations for the example and drew all the diagrams. The financial support of this work by the Department of Defence Production is also gratefully acknowledged.

VI. REFERENCES

BUCHDAHL, H. A. (1954).—" Optical Aberration Coefficients." (Oxford Univ. Press.)
CRUICKSHANK, F. D. (1958).—*Aust. J. Phys.* **11** : 41.
CRUICKSHANK, F. D., and HILLS, G. A. (1960).—*J. Opt. Soc. Amer.* **50** : (In press.)
FORD, P. W. (1959).—*J. Opt. Soc. Amer.* **49** : 875.

THE DAMPING OF WATER WAVES BY SURFACE FILMS

By R. G. Vines*

[*Manuscript received September* 23, 1959]

Summary

Measurements have been made of the stilling of small water ripples by surface films of cetyl alcohol. The damping is considerable and is somewhat in excess of that predicted by existing theories. Imperfections in the ripples induce a directed surface drift (surface mass transport) and it is possible that this is indirectly responsible for the extra damping. Under natural conditions surface films not only impede the formation of small waves but they are also very effective in damping them out.

I. Introduction

The calming action of oil on a water surface has been known from ancient times, but the phenomenon was not studied scientifically until the nineteenth century (see, for example, Aitken 1883). Lamb (1932) examined the theoretical aspects of the damping of water waves and a section of his work was devoted to the influence of surface films on capillary waves. More recently similar treatments were put forward independently by Levich (1941) and Dorrestein (1951).

Experimental measurements by Brown (1936) showed a marked increase in the damping of small ripples on water when surface films were present. However, no quantitative results were reported except for clean water; since the wave frequency was high (~300 c/s) damping must have been pronounced, and with surface films the experimental errors were probably large. Similar measurements have now been carried out at lower frequencies in the presence and absence of cetyl alcohol monolayers. The results are in only fair agreement with existing theoretical treatments, and the observed differences are well beyond the likely experimental error.

II. Experimental

The apparatus used is shown in Figure 1; it was similar to the arrangement described by Brown (1936) in that stroboscopic illumination of the waves was provided by a small vibrating mirror.

Light from a ribbon projection lamp P passed through a cylindrical lens L_1 and a narrow horizontal slit S situated close to the lamp. The concave mirror M_1 (radius 26 cm), rigidly fixed to one arm of a 50 cycle electric tuning fork, was used to produce an image of the slit at the point X, and the light was then rendered parallel by a large lens L_2 separated from X by a distance equal to its focal length (~15 cm). A plane mirror M_2, attached to a travelling telescope T, now deflected the light so that it fell normally upon the water surface in the enamelled tank B; from here it was reflected back vertically into the telescope. The tank B was $\sim$3 cm deep, 30 cm wide, and 35 cm long and it was mounted on sponge rubber to reduce the effect of stray vibrations.

---

* Division of Physical Chemistry, C.S.I.R.O., Chemical Research Laboratories, Melbourne.

Small waves were generated in the tank by the dipper D which just touched the water surface. This dipper was a microscope slide $\sim 7\frac{1}{2}$ cm across, and it was held by a thin wooden rod directly connected to the tuning fork F with a wire (Watson 1901). As a result of this arrangement the frequency of the tuning fork was reduced to 49·4 c/s. When the fork was in operation ripples were obtained on the water which were exactly synchronized with the movement of the mirror M_1.

Through the telescope a series of stationary, narrow lines of light could be seen. These were images of the slit produced by the curvature of the ripples at their troughs and crests. The images from the troughs were above and those from the crests below the water surface, at heights which changed progressively with distance as the amplitude of the ripples decreased (cf. Watson 1901). In the present experiments only the images above the water surface—reflections from the troughs—were followed. The telescope was focused and one of the slit images centred on the cross-wires: a small pointer mounted on a travelling stage was then raised or lowered until it, too, was in focus, and its height above the water surface was measured. The telescope was now moved along an integral number of wave lengths, refocused, and the height of the images above the surface determined again with the pointer. In this way it was possible to observe, indirectly, the change in amplitude of the ripples at varying distances from the dipper, or wave source.

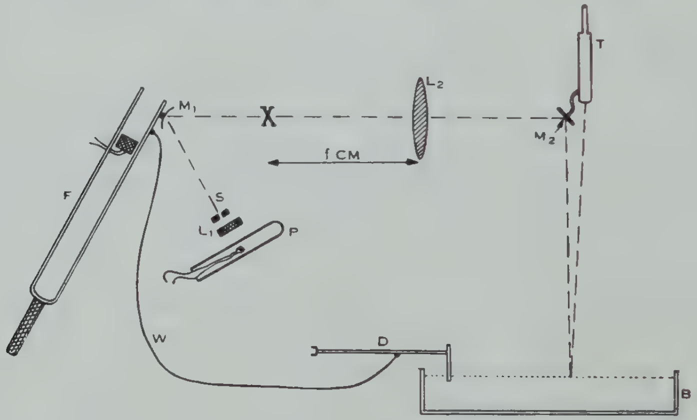


Fig. 1.—Diagram of equipment used. B, tank; D, dipper; F, 50 c/s tuning fork; L_1, cylindrical lens; L_2, lens of focal length f; M_1, M_2, mirrors; P, lamp (18A, 6V); S, slit; T, telescope; W, wire; X, image of slit.

In some experiments the dipper was dispensed with and ripples were generated with a jet of air from a modified fish-tail burner. With care it was possible to stabilize the air stream so that waves of frequency exactly twice that of the tuning fork were obtained: under these circumstances stationary images were again observed through the telescope. It was more difficult to produce air-blown ripples at 49·4 c/s; however, a few rough measurements were made at this frequency.

The wave lengths of the ripples were determined at the same time as the other measurements were made; over the entire distance of travel of the telescope (see Fig. 2) there was no detectable change in wave length.

III. THEORETICAL

The amplitude of damped straight ripples originating from a long dipper is given by

$$a=a_0e^{-\mu x}, \quad \ldots\ldots\ldots\ldots \quad (1)$$

where μ is the damping factor and x is the distance from the wave source at which the ripples possess an initial amplitude a_0. When parallel light falls vertically on water ripples of this type, the troughs act as cylindrical mirrors and series of focal lines are produced above the surface. The focal length f of any one of the mirrors is related to the wavelength λ and amplitude a of the ripple, by the equation (Brown 1936)

$$a=\lambda^2/8\pi^2 f. \quad \ldots\ldots\ldots\ldots \quad (2)$$

Thus, from (1),

$$1/f=(1/f_0)e^{-\mu x}, \quad \ldots\ldots\ldots\ldots \quad (3)$$

where f_0 is the focal length of a ripple of amplitude a_0. It follows that the damping factor μ is given by the slope of the linear plot*

$$\ln f=\mu x+\text{constant}. \quad \ldots\ldots\ldots\ldots \quad (4)$$

Theoretical treatments of damping have been concerned with extended wave trains, where *all* ripples have the same initial amplitude a_0, which everywhere decreases with time t according to the equation

$$a=a_0e^{-t/\tau}, \quad \ldots\ldots\ldots\ldots \quad (5)$$

τ being the modulus of decay.

Although equations (1) and (5) refer to different physical situations, the modulus of decay and the damping factor μ are simply related to each other in terms of the "group velocity" V of the wave train (Dorrestein 1951, p. 272), that is,

$$\tau=1/\mu V. \quad \ldots\ldots\ldots\ldots \quad (6)$$

According to theory (Lamb 1932, p. 624) the modulus of decay τ_1 when viscosity is the controlling factor in damping, is given by

$$\tau_1=1/2\nu k^2: \quad \ldots\ldots\ldots\ldots \quad (7)$$

---

* To obtain a value for μ it is not necessary to know the actual position of the wave source, and this was never recorded in the experiments. All measurements were made at distances greater than ~5 cm from the dipper, because of excessive deformation of the surface in its immediate vicinity.

here ν is the kinematic viscosity and $k=2\pi/\lambda$, λ again being the wavelength. For water at 20 °C ($\nu=0\cdot01$ cm^2/sec) this gives

$$\tau_1 \simeq 1\cdot25\lambda^2. \quad \ldots\ldots\ldots\ldots \quad (8)$$

When surface films are present, the extensions and contractions of the surface associated with the passage of waves are opposed by the variation in surface tension so produced. It is this which is mainly responsible for the damping of the ripples. Under these conditions (Dorrestein 1951)

$$\tau_2=(8/\nu k^2\sigma)^{\frac{1}{2}},$$

where

$$\sigma^2=gk+T'k^3: \quad \ldots\ldots\ldots\ldots \quad (9)$$

here g is the acceleration due to gravity, T' the "specific" surface tension (surface tension/density), and k, as before, is $2\pi/\lambda$. For water the density ρ is unity and T' is numerically equal to the surface tension T, which is derived from the well-known equation for the velocity of propagation v of surface waves under the combined influence of gravity and surface tension (Thomson 1871),

$$v=(g\lambda/2\pi+2\pi T/\lambda\rho)^{\frac{1}{2}}. \quad \ldots\ldots\ldots\ldots \quad (10)$$

This wave velocity v is equal to the product $n\lambda$ where n is the wave frequency.

From equations (**7**) and (**9**) values of μ corresponding to τ_1 and τ_2 may be obtained in terms of equation (**6**). These are:

$$\mu_1=2\nu k^2/V, \quad \ldots\ldots\ldots\ldots \quad (11)$$

$$\mu_2=(\tfrac{1}{8}\nu k^2\sigma)^{\frac{1}{2}}/V. \quad \ldots\ldots\ldots\ldots \quad (12)$$

The value of the group velocity V is calculated from the relation (Lamb 1932)

$$V=v-\lambda \mathrm{d}v/\mathrm{d}\lambda. \quad \ldots\ldots\ldots\ldots \quad (13)$$

IV. RESULTS

(*a*) *Clean Water*

The tank B was thoroughly cleaned and filled with water at $\sim$20 °C and it was allowed to overflow gently to reduce the concentration of impurities at the surface. Typical results are shown in Figure 2 (*a*), where the logarithm of f, the distance of the focal lines above the water surface, is plotted as a function of distance x along the tank (the zero is quite arbitrary and x does *not* represent the exact distance from the dipper). The slope of the line drawn is 0·029/cm and this must be multiplied by 2·303 to obtain the value of μ_{obs} ($0\cdot06_5$ cm). The observed wavelength was 0·595 cm (wave frequency 49·4 c/s) and, as may be seen in Table 1, the corresponding theoretical value of μ_1 (cf. equation (**11**), with $V=40\cdot9$ cm/sec) is 0·055/cm. From the present measurements the surface tension of water, equation (**10**), is 72·3 dyne/cm.

Similar measurements by Brown (1936) give $T=72\cdot1$ dyne/cm for the water used in his experiments. Corresponding to a wave frequency of 300 c/s ($\lambda=0\cdot172$ cm) his value for μ_{obs} is 0·45/cm: the theoretical value (equation (**11**) with $V=77\cdot0$ cm/sec) is 0·35/cm. Theory requires that τ_1, the modulus of decay, be proportional to λ^2, equation (8), and calculation of the proportionality constant gives $1 \cdot 03$ from the present results, and $0 \cdot 97$ from Brown's results. Although these figures are reasonably constant in themselves, they are both about 20 per cent. lower than the theoretical value of $\sim 1 \cdot 25$; thus the damping is somewhat greater than predicted.

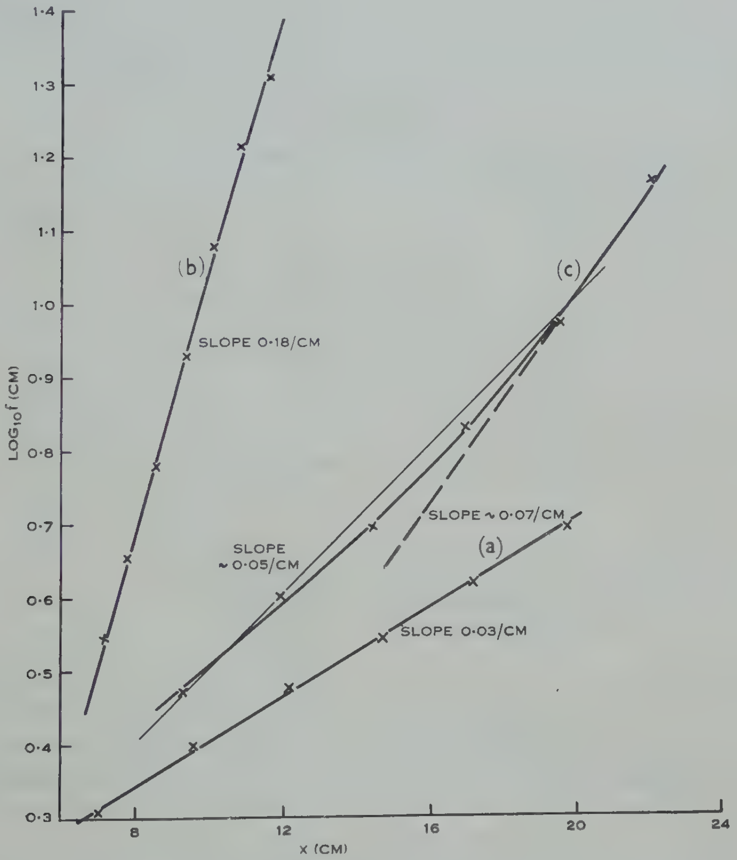


Fig. 2.—The damping of small water ripples as a function of distance x. (a) Clean water surface, (b) with high pressure cetyl alcohol film, (c) with trace of ethyl stearate.

Better agreement with theory was occasionally obtained during the present experiments, but in general the value of μ_{obs} was rather too high. This may have been due in part to the presence of slight traces of surface impurities ; however, in view of the results of subsequent measurements with cetyl alcohol films, it is probably a genuine effect.

(b) *Surface Films*

A surface film was introduced by spreading one drop of a solution of cetyl alcohol in a low boiling petroleum ether. The monolayer obtained was of very high pressure ($\sim$47 dyne/cm) and, as is often the case when spreading takes place from solution, this exceeds the equilibrium pressure of a cetyl alcohol film in contact with the solid. A typical result is shown in Figure 2 (*b*); it may be seen that, in comparison with Figure 2 (*a*), the damping is greatly increased, for the slope of the line has now risen to 0·18/cm ($\mu_{obs}=0\cdot41$/cm when $T=25\cdot5$ dyne/cm). Other measurements were made at different surface pressures, and the results of these and the earlier experiments are summarized in Table 1. In this table μ_1 is the theoretical damping factor as derived from equation (**11**) (viscosity damping) and μ_2 is the corresponding value obtained from equation (**12**) (damping due to variations in surface tension).

TABLE 1

MECHANICALLY GENERATED RIPPLES (WAVE FREQUENCY 49·4 C/S)

	Surface Tension T (dyne/cm)	Wave-length λ (cm)	Slope of $(\log_{10} f)/x$ Graph	V/v (cf. eqn. (**13**))	μ (per cm)			
					μ_1	μ_2	$\mu_1+\mu_2$	μ_{obs}
Clean water	72·3	0·595	0·029	1·39	0·055	—	—	0·065
	72·1	0·172*	$0\cdot19_5$	1·49	0·35	—	—	0·45
Surface films	65·5	0·575	0·11†	1·39	0·06	$0\cdot17_5$	$0\cdot23_5$	$0\cdot25_5$
	58·5	0·555	0·12	$1\cdot38_5$	0·07	$0\cdot18_5$	$0\cdot25_5$	$0\cdot27_5$
	42·5	0·500	$0\cdot13_5$	1·37	$0\cdot09_5$	0·23	$0\cdot32_5$	0·31
	25·5	0·425	0·18	1·35	$0\cdot15_5$	$0\cdot32_5$	0·48	0·41

* Measurements by Brown (1936), at wave frequency 300 c/s.

† This result was obtained at a considerable distance from the dipper: nearer to it the damping was less (cf. Section IV (*d*)).

It may be seen from Table 1 that, when surface films are present, μ_{obs} is distinctly greater than μ_1 or μ_2 in every case: in fact μ_{obs} is roughly equal to the sum of μ_1 and μ_2 over the experimental range. This suggests that damping effects due to bulk viscosity and to surface films are largely independent and that the total damping is a combination of them both. However, such a view receives little support from Dorrestein's treatment, in which damping by surface films is shown to be associated with changes in the nature and extent of hydrodynamic oscillations beneath the surface. Thus the relation $\mu_1+\mu_2\simeq\mu_{obs}$ is probably coincidental.

According to Dorrestein the most general situation is that in which the film, besides showing variation of surface tension with area, also possesses an appreciable surface viscosity; under these conditions the total damping factor is sometimes considerably greater than μ_2. Nevertheless, calculation shows that in the present experiments any such effect is very small indeed,* and it is quite clear that the observed damping is always in excess of the theoretical value, although the discrepancies are never very large. In Table 1 the difference between μ_{obs} and μ_2 is consistently equal to $\sim 0\cdot 08$; this suggests the presence of an extra damping effect which is not included in the theoretical treatment.

---

* Dorrestein indicates that μ may be greater than μ_2 if $N>0\cdot06$, where N is a dimensionless number related to the surface compressibility. The compressibility of cetyl alcohol films has been studied by Nutting and Harkins (1939), and in the present measurements the values of N range from about 3 to 40; however, these values are so large that μ exceeds μ_2 by only a small amount. If the surface viscosity is also taken into account μ is correspondingly reduced, and its final value is almost indistinguishable from μ_2.

(c) *Wind-blown Ripples*

Results for ripples produced by the small air jet are shown in Table 2. The surface tension as derived from equation (**10**) ($\lambda = 0\cdot 535/2$ cm when $n = 98\cdot 8$ c/s) is ~ 28 dyne/cm, corresponding to a film pressure of ~ 45 dyne/cm. This is somewhat lower than the maximum pressure obtained previously, and the air blast is possibly responsible; nevertheless, the film pressure is still high and equal to the equilibrium pressure of solid cetyl alcohol. It will be observed that μ_{obs} is again greater than μ_1 or μ_2. However, as in the previous experiments at high surface pressures, μ_{obs} is rather less than the sum $\mu_1 + \mu_2$, the discrepancies being roughly the same in both cases. Again, as before, $\mu_{obs} - \mu_2 = 0\cdot 08$ for the experiment at $49\cdot 4$ c/s, but the difference is much more pronounced at $98\cdot 8$ c/s: if some extraneous damping effect is responsible it increases markedly with wave frequency.

TABLE 2

WIND-GENERATED RIPPLES: CETYL ALCOHOL FILM AT FULL PRESSURE

Wave Frequency n (c/s)	Surface Tension T (dyne/cm)	Wave-length λ (cm)	Slope of $(\log_{10} f)/x$ Graph	V/v (cf. eqn. (**13**))	μ (per cm)			
					μ_1	μ_2	$\mu_1+\mu_2$	μ_{obs}
$98\cdot 8$	~ 28	$0\cdot 535/2$	$0\cdot 32$	$1\cdot 44$	$0\cdot 29$	$0\cdot 53$	$0\cdot 82$	$0\cdot 74$
$49\cdot 4$	assumed 28	$0\cdot 44$*	$0\cdot 16$	$1\cdot 36$	$0\cdot 13$	$0\cdot 29$	$0\cdot 42$	$0\cdot 37$

* This value may be inaccurate since it was difficult to stabilize the waves in these low frequency measurements. The wavelength was not determined directly, but calculated on the assumption that, as before, $T = 28$ dyne/cm.

No measurements were carried out on clean water since there were traces of oil in the compressed air line, which quickly contaminated the surface.

(d) *Films of Very Low Pressure*

According to Adam (1941), ethyl stearate gives rise to a gaseous film if the surface pressure is less than $\sim 0\cdot 03$ dyne/cm. It was of interest to determine whether films of this kind could exert any damping effect, and a minute trace of ethyl stearate was therefore introduced to a clean water surface on which ripples were generated with the dipper, as before. A small but distinct change was immediately observed and the damping was increased.

The results are shown in Figure 2 (*c*), which, this time, is in the form of a curve, i.e. the damping is more pronounced as the distance from the dipper becomes greater. Similar behaviour, though much less marked, is observed with a low pressure cetyl alcohol film (see Table 1). The effect is probably due to movement of the film by surface mass transport, whereby the concentration of the monolayer is decreased in the vicinity of the dipper and increased towards the far end of the tank. Such movement induced by the waves undoubtedly takes place in any film, but its influence is proportionately less with films of high surface pressure.

For the ethyl stearate film the wavelength observed, ($\lambda=0 \cdot 595$ cm at a wave frequency of 49·4 c/s), is the same as that for clean water (cf. Table 1). This again gives $T=72 \cdot 3$ dyne/cm so that the surface pressure must be very small. Corresponding figures for the damping factor are $\mu_1=0 \cdot 055$/cm (from equation (**11**)) and $\mu_2=0 \cdot 16$/cm (from equation (**12**)). The mean slope of the curve *c* in Figure 2 is $\sim 0 \cdot 05$ ($\mu_{\text{obs}} \simeq 0 \cdot 12$/cm), and that of the asymptotic dotted line is $\sim 0 \cdot 07$ ($\mu_{\text{obs}} \simeq 0 \cdot 16$ cm); in either case $\mu_1+\mu_2>\mu_{\text{obs}}>\mu_1$. It follows that the damping effect is quite pronounced, but rather less than is found with denser films. This is a reasonable result.

One interesting experiment was carried out after a film of unknown impurities had collected overnight on a clean water surface. The film had a definite structure and was quite rigid, and even though its pressure was fairly low ($T \simeq 66$ dyne/cm) the damping was pronounced. After some time damping decreased and the surface layer could be seen breaking up under the continued action of the ripples. This shows that films of high rigidity exert a strong damping effect.

V. CONCLUSIONS

In this work the amplitude of the individual ripples is small, being of the order of 10^{-3} to 10^{-4} cm ;* no results are available for larger waves. Theoretical descriptions of the damping of these capillary ripples are confirmed to only a limited degree, for the observed damping is always somewhat in excess of the predicted values. This suggests that the theories are not completely adequate and that some additional damping mechanism is involved. Circular waves are damped more rapidly than straight waves ; however, any extra damping due to curvature of the wave front must be small in the present experiments. The tank used was not large, but the depth of water was always greater than five wavelengths so that, for ripples of such small amplitude, any effects of depth are scarcely significant. It is possible that damping could be augmented by the general circulation in the bulk of the fluid resulting from surface mass transport (cf. Section IV (*d*)). This arises because the waves are imperfect, and the results of surface movement of this kind are not considered in any theory. All the factors mentioned above lead to an increase in damping, and depend markedly on wave frequency in accord with what is observed in Table 2 ; they may all be involved in the measurements.

The experimental facts are at variance with a suggestion of Dorrestein that the stilling of small ripples by a surface film is primarily due to the action of the film in preventing ripple *formation* rather than to a direct damping effect. He based this view on his result that for capillary ripples of wavelength

---

* From equation (**2**) $a \simeq 10^{-4}$ cm, when f is ~ 10 cm.

0·1–1·7 cm the ratio of the decay moduli τ_1/τ_2 was never very large, varying only from about 1·6 to 4 ; this implied, he believed, that the influence of surface films on the rate of decay of capillary waves was relatively small. However, if the damping is considered as a function of distance rather than time (i.e. if the factor μ in equation (**1**) is thought of, and not the factor τ in equation (**5**)) the effect of a film is more obvious : the ratio μ_2/μ_1 is, in fact, greater than the ratio τ_1/τ_2 for capillary waves, since the decrease in surface tension which follows the introduction of a surface film also produces a corresponding decrease in the wave (and group) velocity. The factor $(\mu_1+\mu_2)/\mu_1$ is greater still, and from the present results this would appear to be a more relevant measure of damping efficiencies.

One additional point should be mentioned. Wavelengths are shortened in the presence of monolayers, and since the shorter waves are damped more easily a further damping is effected indirectly.* It follows that surface films not only help to prevent the formation of small ripples as Dorrestein maintains ; they are also very effective in damping them out.

VI. Acknowledgments

The author wishes to thank Dr. J. A. Barker and Mr. W. W. Mansfield, of the Division of Physical Chemistry, C.S.I.R.O., for their advice and assistance during the course of this work.

VII. References

Adam, N. K. (1941).—" The Physics and Chemistry of Surfaces." 3rd Ed. pp. 43–6. (Oxford Univ. Press.)

Aitken, J. (1883).—*Proc. Roy. Soc. Edinb.* **12** : 56.

Brown, R. C. (1936).—*Proc. Phys. Soc. Lond.* **48** : 312, 323.

Dorrestein, R. (1951).—*Proc. Acad. Sci. Amst.* B **54** : 260, 350.

Lamb, H. (1932).—" Hydrodynamics." 6th Ed. (1953 reprint.) pp. 624–32. (Cambridge Univ. Press.)

Levich, V. (1941).—*Acta Phys.-Chim. URSS* **14** : 307, 321.

Nutting, G. C., and Harkins, W. D. (1939).—*J. Amer. Chem. Soc.* **61** : 1180.

Thomson, W. (1871).—*Phil. Mag.* **42** : 368.

Watson, F. R. (1901).—*Phys. Rev.* **12** : 257.

---

* This was strikingly demonstrated when a fragment of solid cetyl alcohol was placed on a clean water surface : the film pressure rose very slowly and the damping of the ripples increased progressively as their wavelength decreased.

SUNRISE AND ECLIPSE EFFECTS ON THE IONOSPHERE AT BRISBANE

By G. G. Bowman*

[*Manuscript received November* 30, 1959]

Summary

A rotating spaced-loop direction-finding system, located at Brisbane, has been operated to investigate sunrise effects by using pulsed 3·84 Mc/s transmissions, (*a*) at normal incidence and (*b*) at oblique incidence. For oblique-incidence recording the transmitter was located at Armidale (bearing 202° and distant 355 km from Brisbane).

Evidence is presented which suggests the formation (at both E- and F_2-layer levels) of several frontal irregularities, spaced some tens of kilometres apart, extending in directions parallel to the sunrise line and travelling, relative to the Earth, with this line. These fronts pass overhead at Brisbane approximately half-way between the 90 km level and ground level sunrise times.

In the post-sunrise period, F_2-layer spreading on the two-hop trace appears, and the spreading width increases for about 2 hr, suggesting an F_2-layer ripple structure, with increasing ripple amplitude. The post-sunrise sporadic E occurrence suggests frontal irregularities, lying close to the sunrise line direction, soon after sunrise, but indicates a swing in direction as time progresses.

Rotating-loop normal-incidence transmissions were also used to investigate effects due to an eclipse of the Sun, at Brisbane, on April 8, 1959. Post-eclipse E- and F_2-layer frontal irregularities, oriented in directions close to the line representing the end of eclipse in the region of Brisbane, suggest a mechanism operating as the eclipse ends, which is similar to that operating at sunrise.

The possibility, that the ripple structure which produces night-time spread-F at Brisbane is generated at sunrise, is discussed.

I. Introduction

Sunrise effects have been investigated using a rotating spaced-loop direction-finding system located at Brisbane (27·5 °S.; 152·9 °E.). This was capable of determining azimuths-of-arrival for signals received away from the vertical, and it recorded simultaneously, on separate displays, two pulsed 3·84 Mc/s transmissions. One transmitter was located at Armidale (bearing 202° and distant 355 km from Brisbane) and the other on the same site as the rotating-loop system. Detailed descriptions of this system and the nature of the records from it are contained in another paper (Bowman 1960*a*). Ionograms and fixed-frequency $h't$ records have also been used in this analysis.

II. *E*-Layer Sunrise Irregularities

(*a*) *Oblique-incidence Effects*

(i) *Introduction*.—During the period of rotating-loop recording in 1958, oblique-incidence records were taken during 12 sunrise periods in July. Within five of these periods, additional traces, of a character not found at other times of the day, were recorded. These traces occurred in a series, spaced from 50 to 100 km apart. Each trace decreased range at a considerable rate and moved towards a range representing the shortest distance from Armidale to Brisbane via the E layer. These traces were found to persist, sometimes for as long as 30 min. When the records were good enough, azimuths-of-arrival could be read for these signals. The distribution of these azimuths was centred roughly on east. This, and the rapid change of range, suggested an association with a sunrise irregularity moving with the speed of rotation of the Earth. Plate 1, Figure 1, shows the nature of the traces recorded.

---

* Physics Department, University of Queensland, Brisbane.

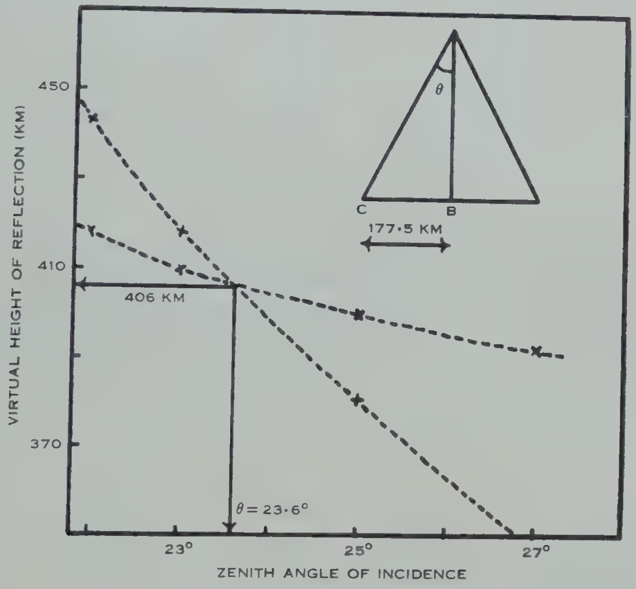


Fig. 1.—Diagram illustrating method for calculation of virtual height of reflection and angle-of-incidence for 3·84 Mc/s oblique-incidence transmissions at 0540 on July 31, 1958.

(ii) *Analysis of Sunrise Effect on July 31, 1958.*—The first step in this analysis was to establish the true ranges of the satellite echoes recorded. To do this, a calculation of the group path of the one-hop F_2-layer signal was necessary, as, because of the method of recording, this has to be used as a reference level. Use is made of the fact that the virtual height of reflection for the equivalent normal-incidence frequency ($f=f' \cos \theta$) is the same as that for the oblique-incidence frequency, f'. The oblique-incidence virtual height of reflection at 0540 has been calculated in the following manner. Table 1 shows, for various angles of incidence θ, (*a*) normal-incidence virtual heights read from the ionogram at the normal-incidence equivalent frequency ($f=f' \cos \theta$) and (*b*) oblique-incidence virtual heights of reflection from the expression, $177 \cdot 5 \cot \theta$ (177·5 km is half the distance between Brisbane and Armidale). Figure 1 shows how a plot of these variations will give (*a*) the angle-of-incidence and (*b*) the virtual height of reflection, at this time. The ionogram is for

Brisbane, and, strictly speaking, the virtual height found (406 km) represents the virtual height of reflection of 3·84 Mc/s transmissions from stations, one midway between Brisbane and Armidale and the other an equal distance on the far side of Brisbane (see Fig. 2 (*a*)). This virtual height gives a group path of 886 km.

TABLE 1

Normal-incidence Equiv. Freq. for Angle-of-Incidence θ (Mc/s)	Normal-incidence Virtual Height of Refln. (km)	Oblique-incidence Virtual Height of Refln. ($177\cdot5 \cot \theta$) (km)
$3\cdot84 \cos 22° = 3\cdot56$	418	443
$3\cdot84 \cos 23° = 3\cdot53$	410	418
$3\cdot84 \cos 25° = 3\cdot48$	400	381
$3\cdot84 \cos 27° = 3\cdot42$	392	348

Similar calculations have been made for other times within this period, and the oblique group paths plotted against time are shown in Figure 2 (*c*) (using the upper time scale). Since the rapid changes in the F_2 layer, around sunrise, travel round the Earth with a speed equal to that of the sunrise line, the propagation parameters for the Armidale to Brisbane reflection point can be taken as the same as those which pertained 2½ min earlier above Brisbane (5 min of time being the longitude difference between Armidale and Brisbane). Thus, in using Figure 2 (*c*) the lower time scale gives the correct group path for the one-hop F_2-layer signal (the *o*-ray has been used for all calculations).

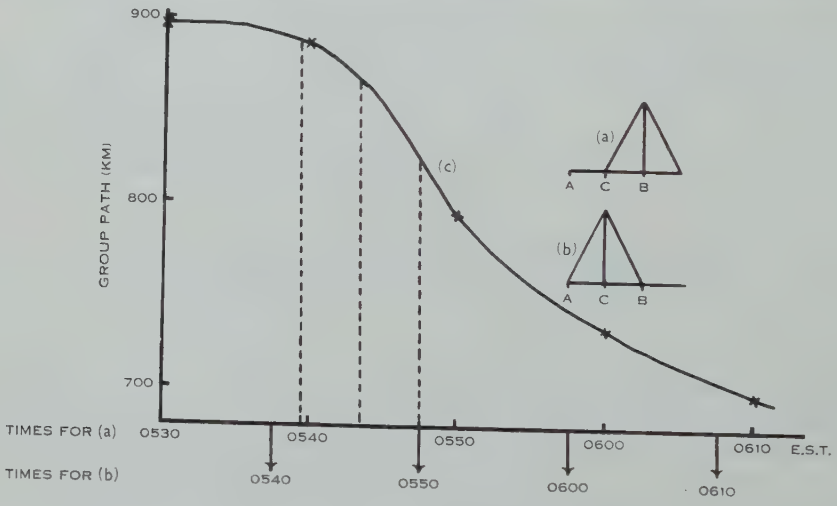


Fig. 2.—Calculated group-path variation of oblique-incidence 3·84 Mc/s transmissions between Armidale and Brisbane on July 31, 1958. *A*, Armidale ; *B*, Brisbane ; *C*, mid point between Armidale and Brisbane.

Table 2 summarizes the observations made on " sunrise " traces of July 31, 1958.

Neglecting group retardation and assuming the reflection height to be 90 km, the data of Table 2 lead to locations, in plan, of the reflection points, shown in Figure 3. The choice of 90 km for the reflection height will be discussed later. In Figure 3, positions are plotted relative to a plane which has a constant orientation with respect to the Sun. Thus, the positions on it of Brisbane and Armidale move with the speed of the sunrise line at 90 km (1500 km/hr). With this arrangement, any ionospheric configuration moving with the sunrise line should appear stationary on the plot. No azimuth information is available for signals recorded at 0550, so loci of possible reflecting points are plotted in this case. The information shown on Figure 3 is consistent with reflection taking place from a series of fronts nearly parallel to the sunrise line and moving with it. The first front passes overhead at Brisbane at about 0601 Eastern Australian Standard Time (E.S.T.).

TABLE 2

Time	Group Path of F_2 Trace (km)	Additional Path of Sunrise Trace (km)	Sunrise Trace Group Path (km)	Sunrise Trace Azimuth
0542	887	275	1162	080°
0546	865	225	1090	085°
		125	990	085°
0550	823	410	1233	—
		350	1173	—
		275	1098	—
		200	1023	—

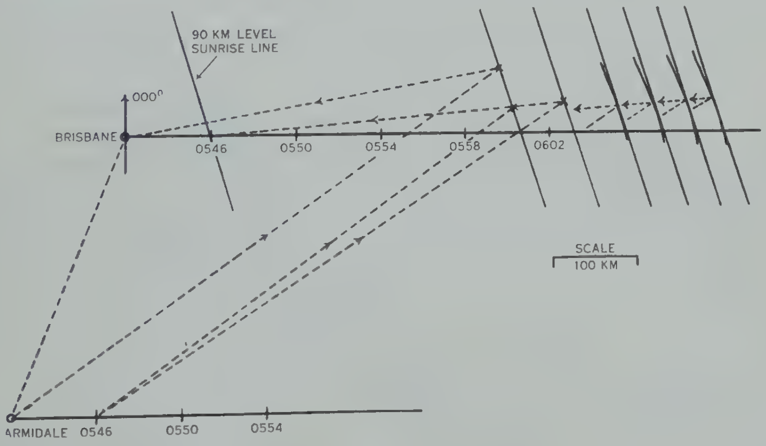


Fig. 3.—Plan position of reflecting surfaces at the 90 km level, around sunrise on July 31, 1958.

(iii) *Analysis of Sunrise Effects on July 13 and 15, 1958.*—A method of calculation similar to that explained in (ii) has been used for the sunrise traces on the oblique-incidence rotating-loop record for July 15, 1958. Figure 4 shows the group path plotted against time for these traces; also for the one-hop F_2-layer *o*-ray trace. Azimuth-of-arrival information is also indicated. The first sunrise trace terminates at a group path corresponding to great circle reflection from the vicinity of the *E* layer. At this point it shows as a well-defined trace with no change of group path for about 1 min. This is, apparently, the smallest range possible for this sunrise trace, and its separation in range from the one-hop F_2-layer trace (*o*-ray) allows a determination of the height of reflection of the sunrise trace. This was found to be (88 ± 5) km. For the plan-position plots (e.g. Fig. 3) the value, 90 km, was assumed. The observed group paths were so large that a few kilometres error in height around this level would not affect the calculated positions of reflection very much.

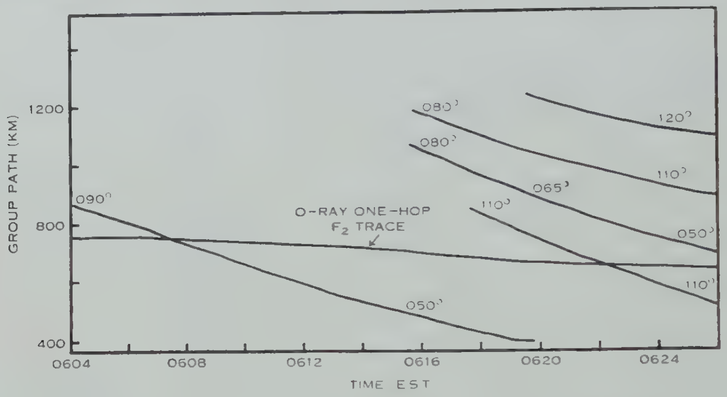


Fig. 4.—Illustration of traces from *E*-layer sunrise irregularities on July 15, 1958.

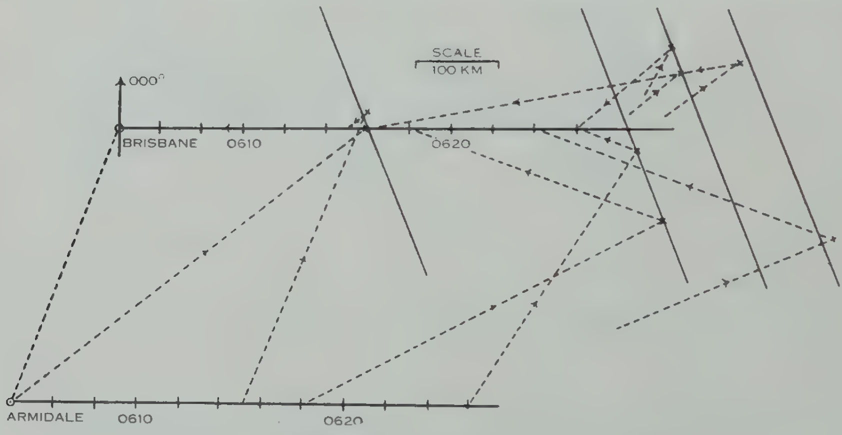


Fig. 5.—Plan position of reflecting surfaces at the 90 km level, around sunrise on July 15, 1958.

Figure 5 is a plot (similar to that of Fig. 3) showing the calculated plan positions of reflecting points determined from the ranges and azimuths shown on Figure 4. In this case the first sunrise front passes overhead at Brisbane at 0616 E.S.T. Of the reflection paths shown in Figure 5, 50 per cent. are such as would be expected if specular reflection occurred from irregularities lying along the sunrise line.

Similar calculations for sunrise traces on July 13, 1958, indicate the initial frontal irregularity passing overhead at 0613 E.S.T., followed by two other reflecting surfaces spaced roughly 100 km apart.

Table 3

Day	Time Transit 1st Irregularity A (E.S.T.)	Ground Sunrise B (E.S.T.)	90 Km Sunrise C (E.S.T.)	B–A (min)	A–C (min)
13.vii.58	0613	0641	0552	28	21
15.vii.58	0616	0640	0551	24	25
31.vii.58	0601	0632	0546	31	15

Table 3 indicates the relationship of times, for the first sunrise irregularity transit, with the times for ground sunrise and 90 km level sunrise, on the three days analysed.

Although normal-incidence recordings were made at the same time as the oblique-incidence records discussed in this section, no traces were seen thereon which could be attributed to reflection from these frontal irregularities.

(*b*) *Post-sunrise Sporadic* E *Orientation*

On certain days in April and May 1959, the rotating-loop system was operated from 0500 to 0800 E.S.T., using the normal-incidence transmitter only. Sporadic E traces were sometimes observed, starting after 90 km sunrise and lasting for several minutes. They showed a definite azimuth-of-arrival. Plate 1, Figure 2, is an example. It has been shown by the writer (Bowman 1960*b*) that sporadic E patches are probably frontal in character. Therefore, these azimuth readings should give an indication of the orientation of the sporadic E fronts. Table 4 shows the days of recording, together with information on these sporadic E occurrences. These short-period sporadic E traces appear to be present on most days.

Baral (1955) has reported, from Calcutta, the occurrence of a " sunrise-time " E_s, appearing generally between E-layer and ground sunrise times. These traces occurred regularly (in more than 60 per cent. of the total observations).

Figure 6 shows a plot of the time separation between this sporadic E occurrence and the 90 km level sunrise time, against the angular separation between the normal to the sporadic E front and the perpendicular to the sunrise line. This reveals that sporadic E fronts occurring soon after the 90 km level sunrise are nearly parallel to the sunrise line. Although the information is meagre, the points on Figure 6 representing sporadic E occurring more than 1 hr after 90 km level sunrise suggest a swing of this orientation as time progresses. This indicates a swing of approximately 90° in $2\frac{1}{2}$ hr.

TABLE 4

Day	Sunrise 90 km Level	Sporadic E Occurrence	Azimuth-of-Arrival	Perpendicular to Sunrise Line	Sporadic E Duration (min)
1959					
April					
8 ..	0526	0538	080°	083°	6
9 ..	0527	—	—	—	—
10 ..	0528	—	—	—	—
11 ..	0528	0800	180°	082°	10
12 ..	0529	0546	100°	082°	3
13 ..	0530	0548	070°	081°	2
16 ..	0533	0600	080°	080°	5
17 ..	0534	—	—	—	—
18 ..	0534	0554	090°	080°	4
19 ..	0535	0642	120°	079°	7
20 ..	0536	0800	170°	079°	12
May					
21 ..	0554	0716	100°	070°	10
22 ..	0555	—	—	—	—
23 ..	0556	—	—	—	—
24 ..	0556	0656	090°	070°	3

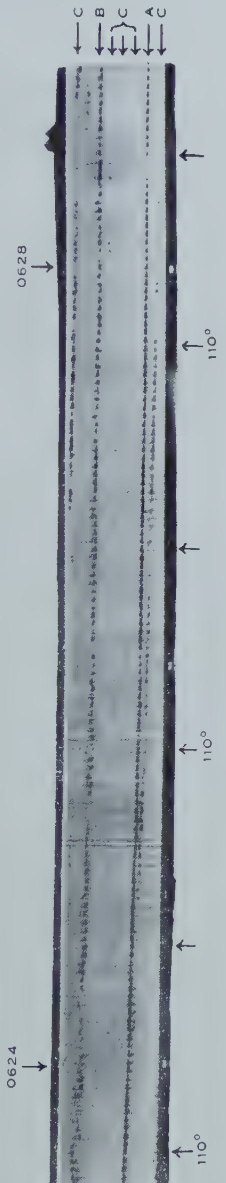
III. F_2-LAYER SUNRISE IRREGULARITIES

(*a*) *Pre-sunrise High Multiples*

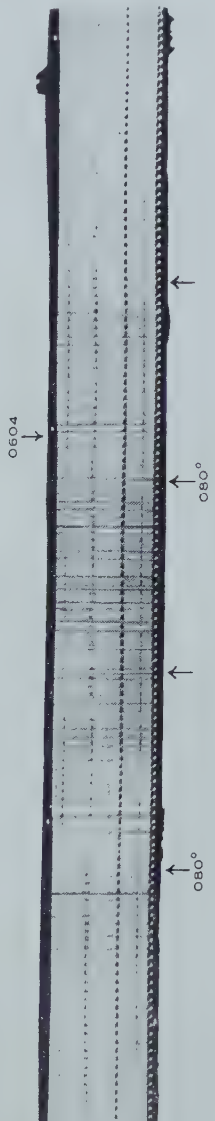
A regular feature of Brisbane ionograms recorded just before sunrise is the presence of high-multiple traces (tenth hop or greater). Plate 2, Figure 2, shows such ionograms and Plate 2, Figure 1, an example on an $h't$ record (at $2 \cdot 28$ Mc/s). The probability that these traces are due to combinations of E-layer and F_2-layer reflections is ruled out, since the spacings between the several high-multiple traces are approximately equal to the one-hop F_2 virtual height. (This is shown most clearly in Figure 7 (*a*) which represents a tracing of part of Plate 2, Figure 1.) The duration of the high multiples which occur at this time is seldom greater than 30 min.

Plate 2, Figure 3, shows a recording of this phenomenon on a rotating-loop record indicating an azimuth-of-arrival of 090°. This is consistent with the irregularity causing the high multiples being frontal in nature and oriented nearly parallel to the sunrise line. The average azimuth-of-arrival for six readings taken in April 1959 was 088°, while the perpendicular to the sunrise line during this period was at approximately 080°. The average azimuth-of-arrival for four readings taken in July 1958 was 082°, and the perpendicular to the sunrise line was at 070°.

SUNRISE AND ECLIPSE EFFECTS ON THE IONOSPHERE AT BRISBANE



1

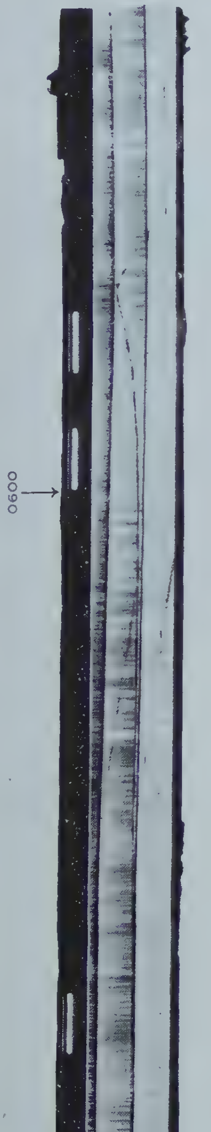


2

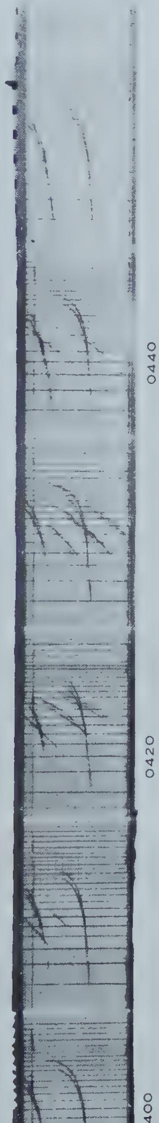
Fig. 1.—Example of oblique-incidence rotating-loop record at 0615 on July 17, 1958, illustrating traces received from sunrise irregularities in the E layer. A, one-hop F_2 trace; B, two-hop F_2 trace; C, traces from sunrise irregularities.

Fig. 2.—Example, on normal-incidence rotating-loop record, of post-sunrise sporadic E occurrence at 0600 on April 16, 1959.

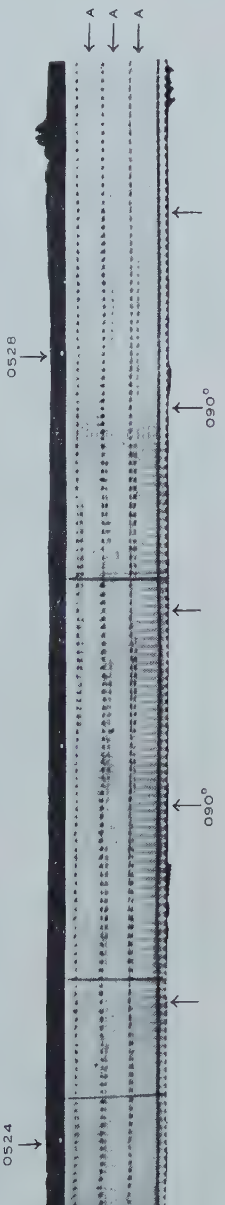
SUNRISE AND ECLIPSE EFFECTS ON THE IONOSPHERE AT BRISBANE



1



2



3

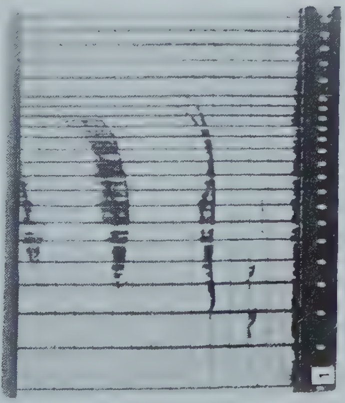
Examples of pre-sunrise high-multiple traces from F_2-layer irregularities.

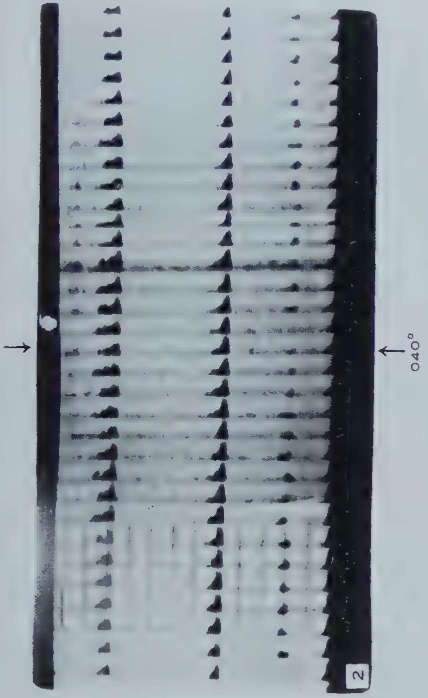
Fig. 1.—2·28 Mc/s $h't$ record on July 27, 1958.

Fig. 2.—Ionograms on September 27, 1956.

Fig. 3.—Normal-incidence rotating-loop record on April 19, 1959, illustrating azimuth-of-arrival of pre-sunrise high-multiple traces. *A*, High-multiple traces.

SUNRISE AND ECLIPSE EFFECTS ON THE IONOSPHERE AT BRISBANE





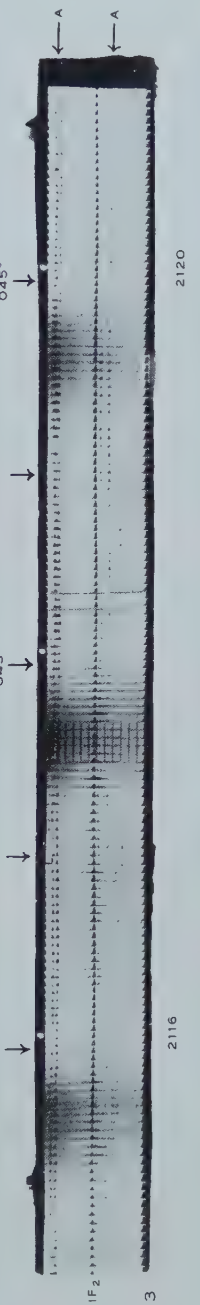


Fig. 1.—Ionogram at 0720 on March 29, 1956, illustrating post-sunrise spread-F on two-hop F_2 trace.

Fig. 2.—Post-eclipse normal-incidence rotating-loop record at 1756 on April 8, 1959, indicating azimuths-of-arrival for sporadic E and spread-$2F$ traces.

Fig. 3.—Post-eclipse normal-incidence rotating-loop record on April 8, 1959, indicating azimuths-of-arrival, and nature of high-multiple traces. *A*, High-multiple traces.

A detailed study has been made of the occurrence of this phenomenon on Brisbane ionograms during 1956. Figure 8 indicates the diurnal variation of percentage occurrence, month by month. Also plotted (to the nearest 10 min) are the times for ground sunrise. The time of occurrence of the pre-sunrise peak in the distribution appears to move with the time of sunrise through the year. This pre-sunrise peak is evident from the results of Baird (1954). Figure 9 shows the aggregate of these distributions plotted relative to the time of ground sunrise. This shows a peak in the distribution approximately 30 min after the 250 km level sunrise, and 50 min before ground sunrise. The average percentage occurrence of 37 per cent. for this peak indicates that, in 1956, this phenomenon occurred regularly. The seasonal variation of percentage occurrence of this phenomenon in 1956 is shown in Figure 10, when occurrence is considered in the period between the time of ground sunrise and an hour prior to this time.

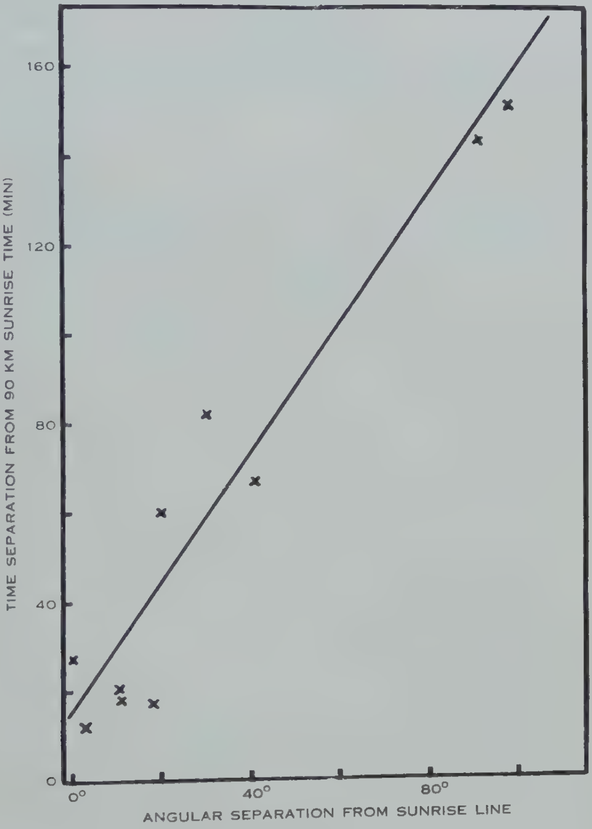


Fig. 6.—Graph of post-sunrise sporadic *E* occurrence. Time separation from 90 km level sunrise time is plotted against angular separation between the normal to the sporadic *E* front and the perpendicular to the sunrise line.

Referring again to Figure 7 (*a*), since the pulse repetition frequency used was 50 c/s, the position of the ground pulse represents 2998 km virtual range when

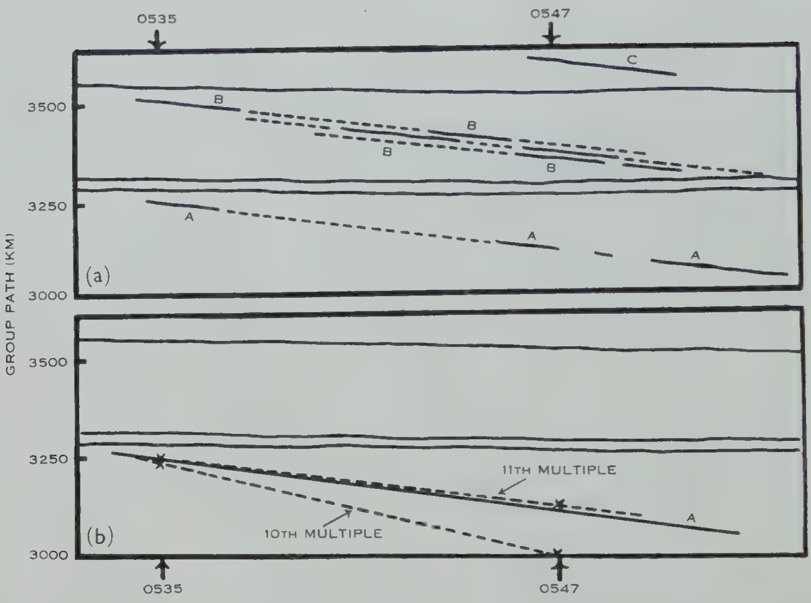


Fig. 7 (*a*).—Tracing of part of Plate 2, Figure 1, illustrating the nature of high-multiple traces in detail.

Fig. 7 (*b*).—Comparison between rate of change of group path for trace *A* from (*a*) and calculated changes for 10-hop and 11-hop transmissions.

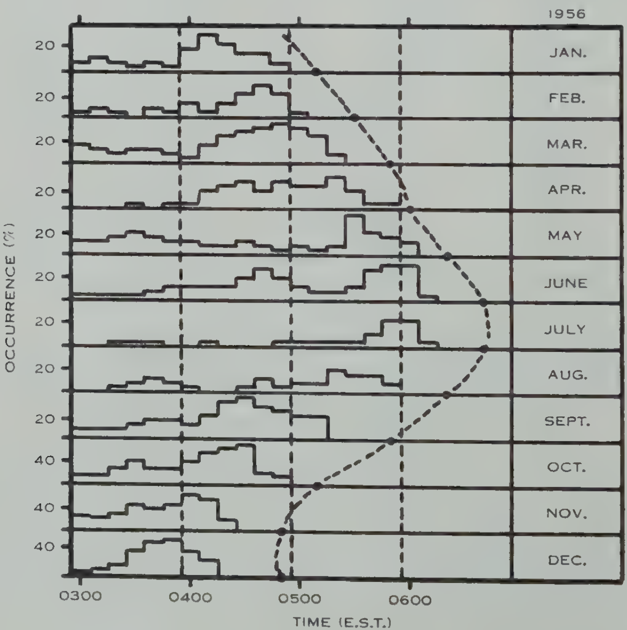


Fig. 8.—Percentage occurrence of high-multiple traces, around sunrise, for the 12 months of 1956. Times of ground sunrise (to the nearest 10 min) are indicated.

considering high-multiple traces. The traces marked *A* appear to correspond with the same irregularity, since they are in line, whereas the traces marked *B* suggest a sequence of irregularities, the reflection point jumping from one irregularity to the other as time progresses. This multiplicity of traces is a common feature of the phenomenon. It is also illustrated by Plate 2, Figures 2 and 3.

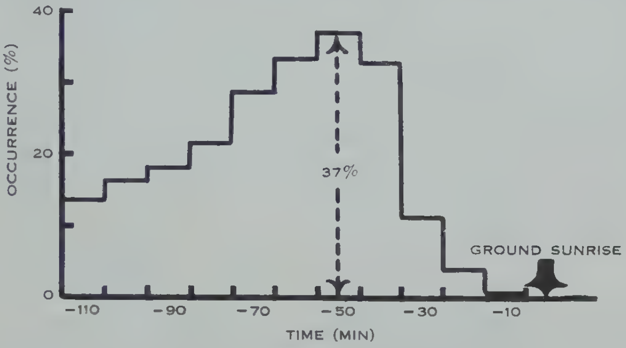


Fig. 9.—Percentage occurrence of high-multiple traces (from ionograms) for 1956 when plotted relative to ground sunrise.

It is possible, in some cases, to estimate which multiple hop the trace represents and the position of the sunrise F_2 irregularity relative to the station at the time of recording, by the following method. Using the information from ionograms, true heights of reflection for a fixed frequency are calculated (Schmerling 1958) at 10-min intervals. It is then assumed that the changes in height arose from the movement of a fixed ionospheric configuration over the Earth with the same velocity as that of the F_2 sunrise line; and a true height versus distance diagram is constructed on this basis. The results for sunrise, July 27, 1958, are shown in Figures 11 (*a*) and 11 (*b*) and the group retardation

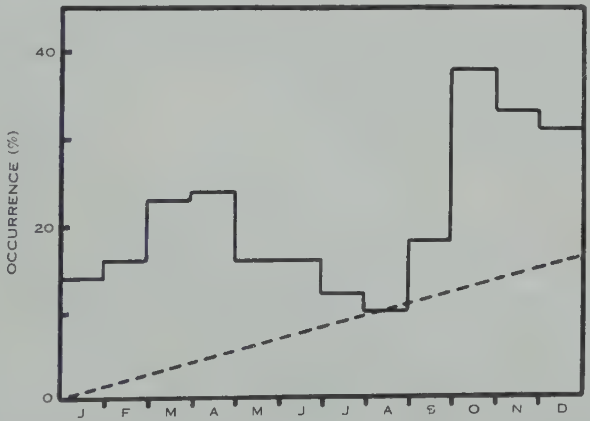


Fig. 10.—Seasonal distribution (for 1956) of percentage occurrence of high-multiple traces for the hour before ground sunrise.

for vertical reflections is shown in Figure 11 (*c*). The frequency used here was 2·28 Mc/s.

It is now possible to determine, by trial, the location of a scattering irregularity on the configuration such that the group path for the multiple-hop echo, at a selected time, would fit the observed results. This is not, however, a simple geometrical problem, since allowance must be made for group retardation at each reflection from the ionosphere. Since this retardation is relatively small (in terms of equivalent path), it is sufficiently accurate to assume it the same for oblique as for vertical reflection. A guess must be made as to the number of hops involved. In the example used (trace *A* at 0535 on July 27, 1958, Fig. 7 (*a*)), since the total group path for trace *A* is just less than 12 times the group path for one-hop F_2, it is clear that the number of reflections involved must be 11 or 10.

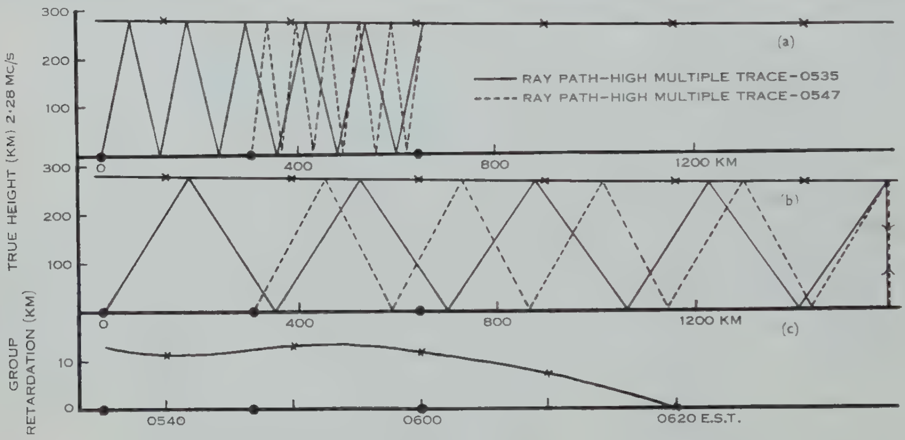


Fig. 11 (*a*).—Proposed ray paths for 11-hop F_2-layer transmissions for high-multiple traces observed at sunrise on July 27, 1958.

Fig. 11 (*b*).—Proposed ray paths for 10-hop F_2-layer transmissions for high-multiple traces observed at sunrise on July 27, 1958.

Fig. 11 (*c*).—Variation of group retardation for vertical-incidence 2·28 Mc/s transmissions at sunrise on July 27, 1958.

Solutions, based on these alternative assumptions, are shown in Figures 11 (*a*) and 11 (*b*), the positions of Brisbane, relative to the configuration, at the selected times, being shown by dots on the horizontal axis. Another time, several minutes later (0547), was then chosen, and the group paths calculated (on the same basis as before) to the position of the irregularity for 10- and 11-hop transmissions at 0535. Figure 7 (*b*) shows the rate of change of group path, calculated in this way, for the alternative transmissions. The rate of change of group path for 11-hop transmission is very close to that observed for trace *A*, and therefore Figure 11 (*a*) must indicate the approximate position of the principal F_2 irregularity. This appears to pass overhead at Brisbane at approximately 0600. This time is 34 min before ground sunrise.

(*b*) *Post-sunrise Two-hop Spread*-F

Another frequently recurring feature of ionograms at Brisbane is the presence of unresolved satellite traces related to the two-hop F_2 trace. This type of spreading commences a short time (within 30 min) after ground sunrise. Plate 3, Figure 1, illustrates an example of this. Calculations have been made of the average spreading widths (in range) of these spread-$2F$ traces for March 1956, April 1956, and July 1958. These average spreading widths are shown, by Figure 12, to grow from a value slightly greater than the transmitter pulse width of 10 km to a width in the region of 40 km about 2 hr later. After this, there is a gradual fall in average spreading width, at least until 1200 (no readings were taken after 1200). The times of commencement of this phenomenon and the times of maximum average spreading widths appear to be closely related to the time of ground sunrise (Fig. 12).

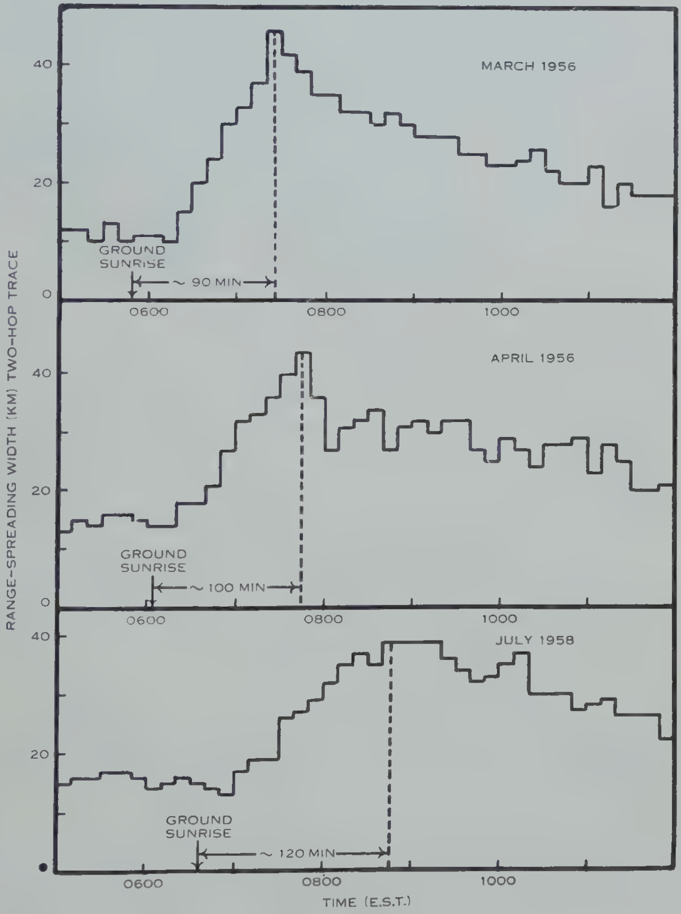


Fig. 12.—Variation of average range-spreading width of two-hop spread-F between 0500 and 1200 for March 1956, April 1956, and July 1958.

In view of the apparent orientation of the sunrise F_2 irregularities along directions close to the sunrise line (reported in Section III (*a*)), and also the sunrise line orientation of sporadic E and its subsequent swing in direction

(Fig. 6), the azimuths-of-arrival of these satellite echoes appear to be of interest. Attempts were made during April and May 1959 to measure the azimuths-of-arrival on the assumption that these ionospheric irregularities were of a nature similar to those causing spread-F at night (Bowman 1960*a*). However, the signal strength of these echoes was so low that they were not recorded by the rotating-loop system.

IV. Post-eclipse Ionospheric Effects

During and after the eclipse of the Sun at Brisbane on April 8, 1959, normal-incidence rotating-loop records were made. During the eclipse nothing of interest concerning the azimuth-of-arrival of signals was recorded.

However, from 1600, about half-an-hour after the end of the eclipse, two-hop F_2-layer spreading appeared, the spreading width gradually increasing until 1830 when it was 150 km, made up of some 10 or more satellite traces. A plot of the variation of spreading width with time is shown in Figure 13. After 1830, the range of the main two-hop trace was so great that the extent of the spreading could not be determined because of the limited range of recording. It was still present at the end of the rotating-loop recording at 2122. The azimuth-of-arrival of these satellite signals did not vary by more than 20° in this period, ranging from 030° to 050°. Plate 3, Figure 2, illustrates the null at 1756. Figure 14 indicates that the spread-F fronts, in this case, are aligned roughly in the same direction as the line representing the end of the eclipse at Brisbane at 1527 E.S.T. Under normal conditions this orientation of the spread-F fronts is rare (Bowman 1960*a*).

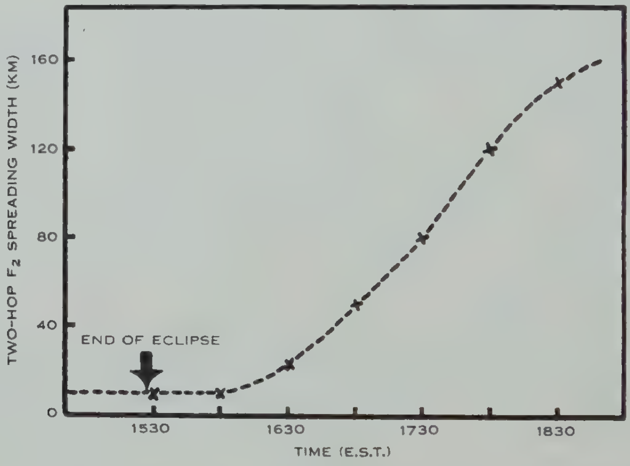


Fig. 13.—Variation of range-spreading width of two-hop spread-F at 3·84 Mc/s for 3 hr after the end of the eclipse at Brisbane on April 8, 1959.

No nulls were found in the sporadic E traces present from 1530 until 1550 and from 1616 until 1644, but the trace present from 1740 until 1820 (illustrated in Plate 3, Fig. 2) gave an azimuth-of-arrival of approximately 040°.

From 2012 until the end of recording at 2122, high-multiple traces were present on the rotating-loop record, indicating echo ranges between 3000 and 3600 km. These traces showed little change of range during this time, but they did show a change in appearance. They first appeared as clean traces, but by 2122 they showed a number of satellite traces. The azimuth-of-arrival of the high-multiple traces was consistent with that found for two-hop spread-F and sporadic E in this period (between 030° and 050°). Plate 3, Figure 3, illustrates the nature of high-multiple traces at 2120.

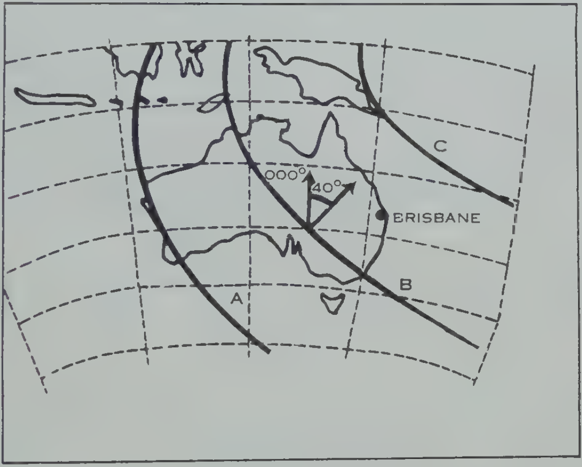


Fig. 14.—Illustration of times for the end of the Sun's eclipse over Australia on April 8, 1959. *A*, Eclipse ends 1400 E.S.T. ; *B*, eclipse ends 1500 E.S.T. ; *C*, eclipse ends 1600 E.S.T.

V. DISCUSSION

The results of this present investigation suggest that, at a time approximately half-way between 90 km sunrise and ground sunrise, a series of frontal irregularities, having spacings of the order of tens of kilometres, are created at both the E- and F_2-layer levels. These have been detected in the E layer by oblique-incidence recording, and in the F_2 layer by high-multiple traces which appear on normal-incidence records about an hour before ground sunrise. These frontal irregularities (E and F_2 layer) are shown to be oriented close to directions parallel to the sunrise line and appear to move with it.

Evidence is presented to show that, after the occurrence of these irregularities (moving with the sunrise line) in the E and F_2 layers, irregularities of the same form appear to remain in both layers. Figure 12 shows that, if the F_2-layer irregularities be regarded as " ripples " in the ionization contours, the amplitude of these ripples probably increases after sunrise to a maximum from $1\frac{1}{2}$ to 2 hr later, since the spreading width appears likely to be a function of ripple amplitude. (The conditions for the occurrence of two-hop spreading, when one-hop spreading is absent, are discussed by Bowman (1960*a*).) Evidence for post-sunrise sporadic E irregularities, aligned close to the sunrise line, is obtained from the small patches of sporadic *E* recorded soon after the sunrise period (Fig. 6). Baral (1955) has reported isolated sporadic *E* occurrence, mainly during the sunrise period.

Bracewell *et al.* (1951) and Lauter and Schmelovsky (1958) have detected sunrise effects on very-low frequency transmissions, which may possibly be related to the results of this paper. Recording the signal strength of 16 kc/s oblique-incidence transmissions, Lauter and Schmelovsky (1958) found a sudden reduction (of the order of 20 dB) in signal strength at the time for 90 km level sunrise, at the mid point of the transmission path. A minimum of signal strength occurred between 90 km level sunrise and ground sunrise (for the mid point of the transmission path). This minimum was followed by a broad peak in signal strength (about ground sunrise), after which it assumed a normal day-time value.

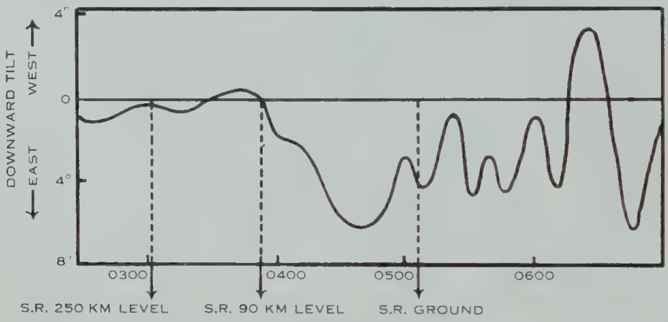


Fig. 15.—Variation of F_2 layer tilt around sunrise at Slough (after Bramley 1953).

It is interesting to note, that, in the case investigated, of pre-sunrise F_2-layer high multiples, the F_2-layer sunrise irregularity appears to pass overhead at approximately the same time (0600 on July 27, 1958) as the leading *E*-layer sunrise irregularity four days later (0601 on July 31, 1958). This suggests an effect generated at the *E*-layer level and transmitted to the F_2-layer by electrodynamic coupling.

Figure 15 reproduces a result due to Bramley (1953) showing the variation of the F_2-layer tilt around the sunrise period. Times for sunrise at 250 km, 90 km, and ground levels are indicated. Figure 15 reveals a tilt of the F_2 layer which increases with time and commences at the 90 km level sunrise rather than at the 250 km level sunrise. This may be important in view of the coupling suggested earlier. As pointed out previously, Figure 12 suggests ripples of increasing amplitude after sunrise. Therefore the marked oscillation of the F_2-layer angle of tilt after the initial sunrise effect (Fig. 15) appears to be consistent with the results shown in Figure 12.

Conditions in the ionosphere after the eclipse on April 8, 1959 are very similar to those found after sunrise. If the uncovering of the Sun after the eclipse be regarded as a virtual sunrise, the line along which the eclipse ends is analogous with the sunrise line. The gradual increase in two-hop F_2 spreading widths after the eclipse, and the continued orientation of *E*- and F_2-layer irregularities along directions close to the line representing the end of the eclipse, all point to a mechanism operating as the eclipse ceases that is comparable with that operating at sunrise.

The possibility arises that ripples, generated either at sunrise or sunset, are responsible for the one-hop spread-*F* observed at night at Brisbane. Effects, similar to those discussed here for sunrise, have not been detected at sunset, but there is some evidence to suggest that ripples, generated at sunrise, persist throughout the day, and are responsible for the spread-*F* observed at night. The two-hop, post-sunrise spreading width, illustrated by Figure 12, should be a measure of the amplitude of the ripples. At 1200, it is obvious that a ripple structure is still present, particularly for July, a month of maximum spread-*F* occurrence. Another possible association between ground sunrise and night-time spread-*F* is shown in Figure 16. Here, the seasonal distribution of spread-*F* (for the years 1944–1958 inclusive) is compared with the seasonal variation in the angular separation between the sunrise line and the magnetic meridian. The similarity is interesting, particularly as the spread-*F* minima are displaced somewhat from the equinoxes and occur roughly at times during the year when the sunrise line lies along the magnetic meridian.

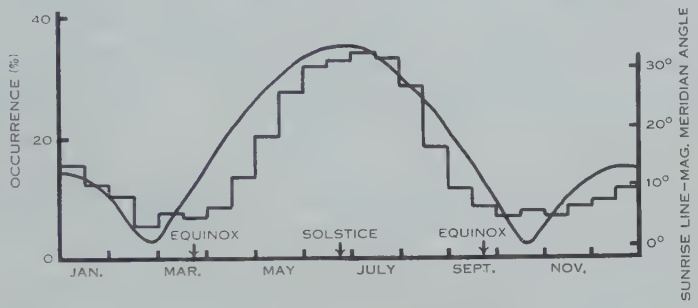


Fig. 16.—Diagram illustrating comparison between seasonal variation of spread-*F* at Brisbane (1944–1958 inclusive) and the angular separation between the sunrise line and the magnetic meridian.

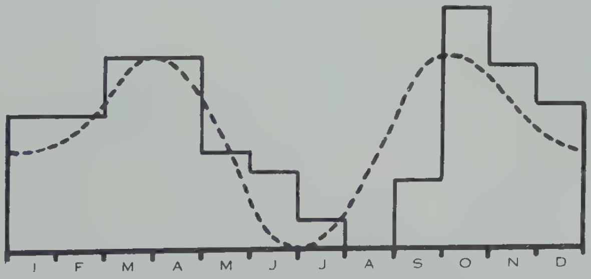


Fig. 17.—Seasonal distribution of sunrise high multiples for 1956, after an allowance is made for a possible sunspot-cycle variation.

Another apparent association between sunrise and night-time phenomena is revealed by considering the seasonal distribution of the sunrise high multiples.

During 1956 the sunspot activity rose sharply, and therefore the distribution shown in Figure 10 suggests a direct relationship with sunspot activity. If this association is assumed, a seasonal distribution, with some independence from sunspot activity, is obtained by subtracting from the percentage occurrence of each month, the background variation proposed on Figure 10. Figure 17 shows the distribution deduced in this way, revealing equinoctial maxima, a winter minimum, and a summer sub-minimum. This distribution is the inverse of that found for spread-F at Brisbane. One possible clue to this inverse relationship is the need for a small ripple amplitude to produce high multiples (Baird 1954) and much larger amplitudes for one-hop spread-F.

VI. Acknowledgments

The author would like to thank Professor H. C. Webster for his continued guidance; also Dr. R. W. E. McNicol and Dr. J. A. Thomas for their assistance during the course of the work. The ionograms were kindly supplied by the Ionospheric Prediction Service. This work was financially supported by the Radio Research Board of C.S.I.R.O., and was carried out while the author was the holder of a C.S.I.R.O. studentship.

VII. References

Baird, K. (1954).—*Aust. J. Phys.* **7**: 165.
Baral, S. S. (1955).—*J. Atmos. Terr. Phys.* **6**: 160.
Bowman, G. G. (1960*a*).—*Planet. Space Sci.* **2**. (In press.)
Bowman, G. G. (1960*b*).—*Planet. Space Sci.* **2**. (In press.)
Bracewell, R. N., Budden, K. G., Ratcliffe, J. A., Straker, T. W., and Weekes, K. (1951).—*Proc. Instn. Elect. Engrs.* III **98**: 221.
Bramley, E. N. (1953).—*Proc. Roy. Soc.* A **220**: 39.
Lauter, E. A., and Schmelovsky, K. H. (1958).—*Beitr. Geophys.* **67**: 218.
Schmerling, E. R. (1958).—*J. Atmos. Terr. Phys.* **12**: 8.

A RELATIONSHIP BETWEEN SPREAD–F AND THE HEIGHT OF THE F_2 IONOSPHERIC LAYER

By G. G. Bowman*

[*Manuscript received November* 30, 1959]

Summary

The variation, throughout the night, of range-spreading spread-F widths, at Brisbane, is investigated. Evidence is presented to support the hypothesis that the ripple amplitude of the spread-F ionospheric irregularities varies directly with the layer height. This leads to an association between the range-spreading width and the height of the layer.

Certain aspects of frequency-spreading are also discussed.

I. Introduction

In a previous paper on spread-F at Brisbane (Bowman 1960,† subsequently referred to as I) the diurnal variation of spread-F was considered statistically, the existence or not of spread-F traces being taken as a measure of its occurrence. No account was made of the widths (in range) of the spread-F traces. Treated in this way, the diurnal variations suggested increased spreading occurrence when the F_2 layer increased in height. It was postulated that, possibly, this resulted from an increased amplitude for the spread-F " ripples ", as the height of the layer became greater. In this case, it would be expected that the width of range spreading, at any particular time, should also be related to the height of the layer, for this spreading width is determined by the range of zenith angles from which satellite signals are possible, and this range should increase with ripple amplitude, as indicated in Figure 1. The present paper deals with this aspect of the phenomenon.

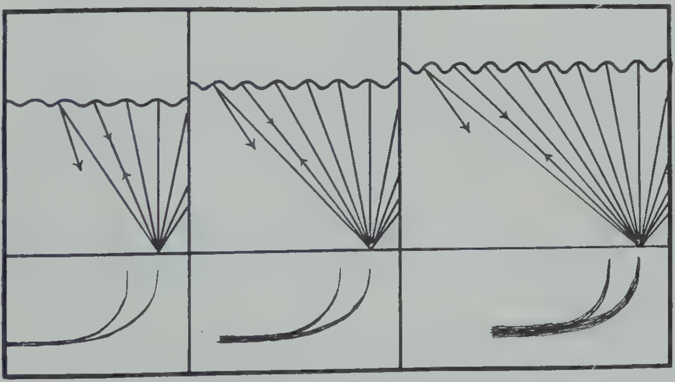


Fig. 1.—Diagram illustrating suggested dependence of range-spreading width on the amplitude of the irregularity ripples.

---

* Physics Department, University of Queensland, Brisbane.

† Bowman, G. G. (1960).—*Planet. Space Sci.* 2. (In press.)

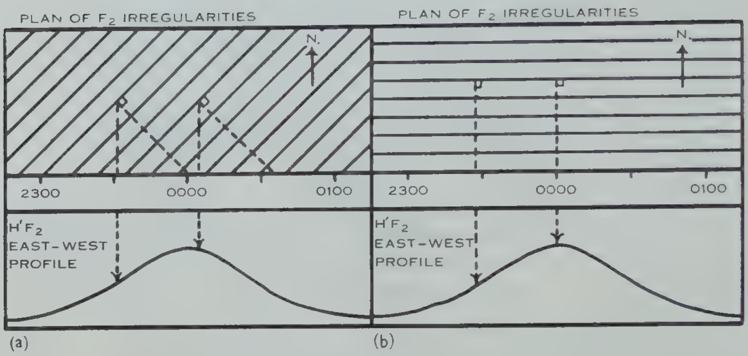


Fig. 2.—East-west profile of height changes and plan position of F_2 irregularities, (*a*) leading to a maximum range-spreading width some time later than time of maximum layer height, (*b*) leading to a maximum range-spreading width at the time of maximum layer height.

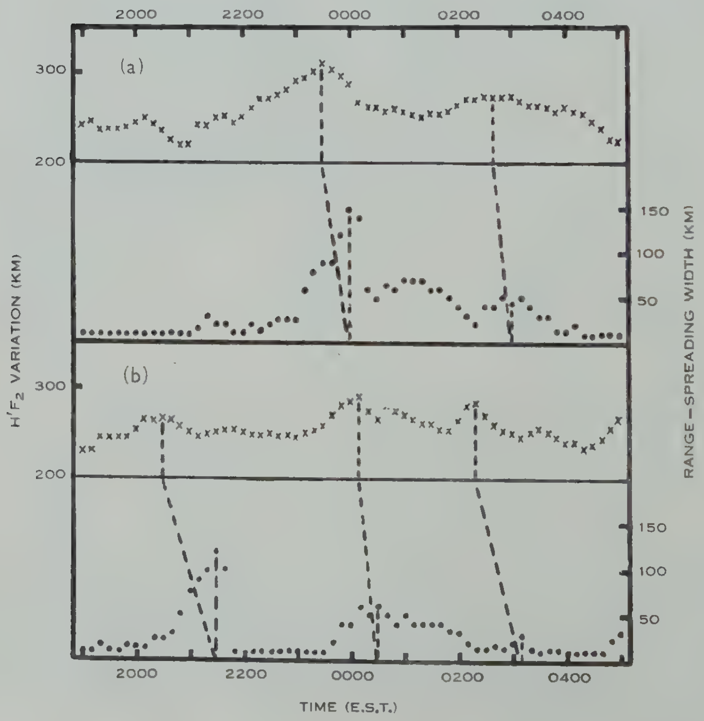


Fig. 3.—A comparison between $h'F_2$ variation and range-spreading spread-*F* widths, (*a*) for the night of August 16/17, 1956, and (*b*) for the night of July 23/24, 1956.

II. Analysis and Discussion

If the hypothesis discussed in Section I is correct, it would be expected that a relationship could be established between height changes and the relative widths of the range-spreading trace. However, this will be complicated by the fact that, in general, the frontal orientation of the ripples is not the same direction as that of the gradient of height changes. In I it was established that the most favoured azimuth-of-arrival for spread-*F* echoes occurred in the north-west quadrant. Therefore, if it is assumed that height changes have an east-west gradient, and the orientation of the ripples is at an angle of 45° to this direction, the arrangement is as illustrated in Figure 2 (*a*). In this figure the height profile is imagined as remaining stationary, while the position of the station moves. The frontal irregularities are closely spaced, and, as an approximation, they are considered to have infinite length. This figure illustrates that, since the ripples, which are responsible for the extreme edge of the range spreading, are appreciably displaced from the zenith, they will correspond to the layer height over the station some minutes earlier. Thus maximum spreading width can be expected some minutes after the maximum layer height. This can usually be observed in the records. Figures 3 (*a*) and 3 (*b*) show examples. For the F_2-layer height the quantity $h_{\min}$ is used.

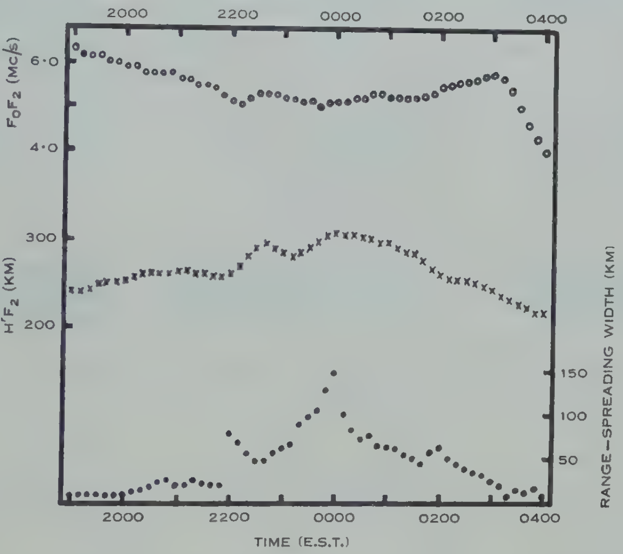


Fig. 4.—A comparison between f_0F_2 variation, $h'F_2$ variation, and range-spreading spread-*F* widths for the night of May 3/4, 1956.

The nature of the spread-*F* phenomenon at Brisbane for the night of May 3/4, 1956 is such as to suggest that some of the complicating parameters are absent. The layer height does not oscillate throughout the night, as it usually does, but shows a more or less gradual rise to midnight, followed by a gradual fall. The peak in range-spreading width coincides with maximum height of the layer, so possibly the frontal orientation of the ripples, at this time, lies in the same direction as the height change gradient. Figure 2 (*b*) shows the suggested configuration. Figure 4 illustrates the f_0F_2, $h'F_2$, and range-spreading width variations on this night, the readings being taken from the ionograms made at 10-min intervals. The range-spreading width variation follows closely the $h'F_2$ variation.

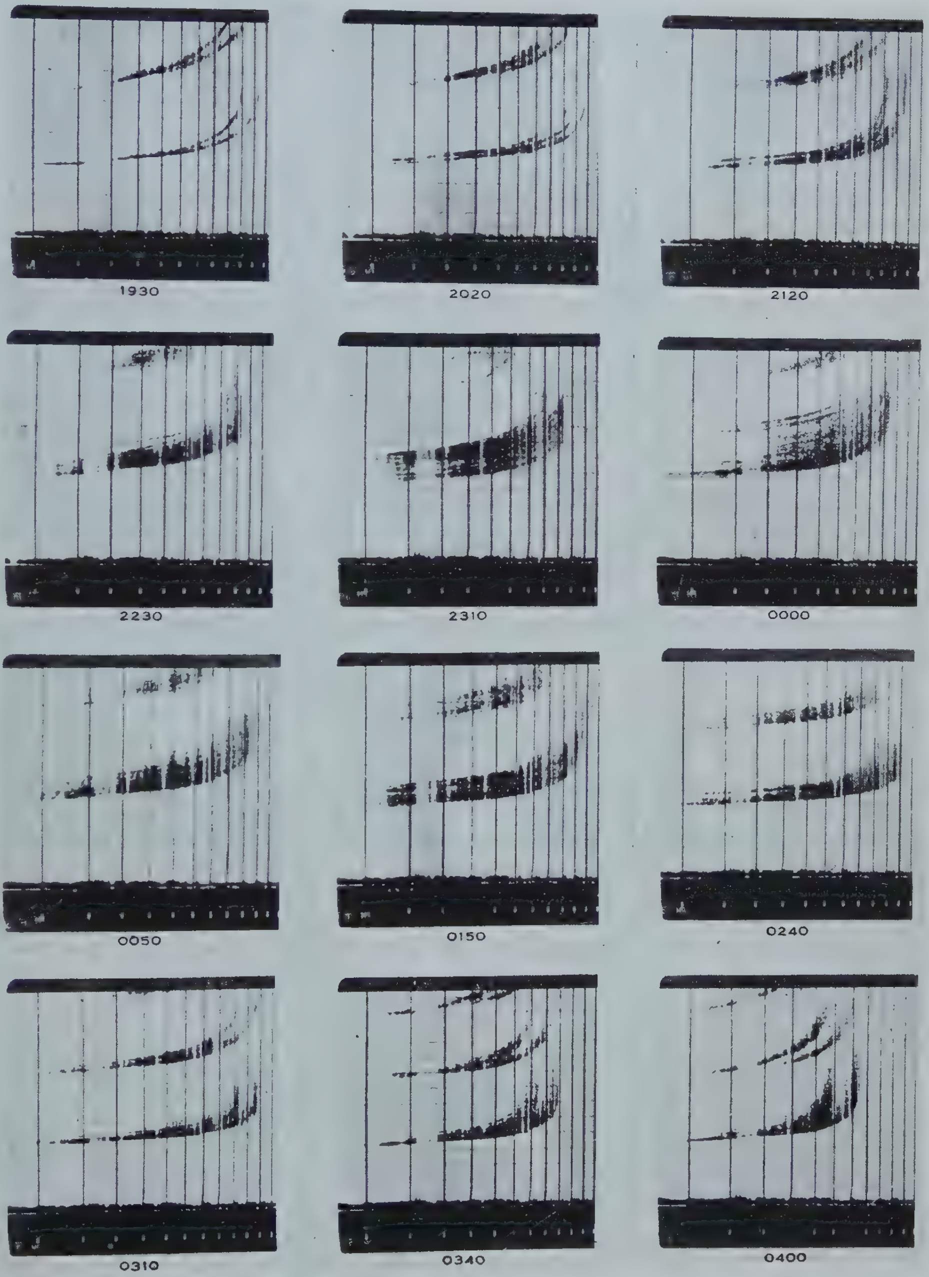
It was explained in I that, possibly, when the ripple amplitudes become relatively small, the ripples can still be detected in the form of frequency spreading, provided there is a sufficiently large f_0F_2 gradient across the ripple structure (cf. I for definitions of range spreading and frequency spreading). The ionograms on this night are particularly clear in showing the transition from range spreading to frequency spreading. It is seen that, by 0300 hr, the range-spreading width has been reduced to a low value, but after this time the ionograms show increasing frequency spreading. Also, after 0300 the f_0F_2 value changes at a fast rate. This appears to be evidence for the mechanism, proposed in I, to explain frequency spreading.

Sample ionograms at various times throughout the night of May 3/4, 1956 are shown in Plate 1.

III. ACKNOWLEDGMENTS

The author would like to thank Professor H. C. Webster for helpful advice in the preparation of this paper. The Ionospheric Prediction Service kindly supplied the ionograms. The work was financially supported by the Radio Research Board of C.S.I.R.O., and was carried out while the author was the holder of a C.S.I.R.O. Studentship.

SPREAD-F AND THE HEIGHT OF THE F_2 LAYER



1930 2020 2120

2230 2310 0000

0050 0150 0240

0310 0340 0400

Several ionograms on the night of May 3/4, 1956, illustrating the varying nature of spread-*F* traces.

THERMAL TRANSIENTS ASSOCIATED WITH NATURAL CONVECTION

By R. C. L. Bosworth* and C. M. Groden†

[*Manuscript received November 6, 1959*]

Summary

An electrical circuit, equivalent to the thermal processes concerned in the act of setting up a system of natural convection in a fluid, has been proposed. This circuit consists of parallel elements, one attributed to convection alone and one to conduction alone. The former consists of a resistance and an inductance in series, the latter of a resistance and a shunt capacity. The nature of the transients associated with each circuit has been derived by Heaviside analysis. Functions, at the most involving no more than two independent variables, are found connecting easily measured characteristics of the transients and the circuit parameters. These functions are, it is proposed, to be used to examine experimental thermal transients and through them to examine their circuit parameters.

I. Introduction

The way in which the distribution of temperature settles down after a sudden thermal shock is ordinarily described by the Fourier equation. In the Fourier equation a single characteristic property of the system is referred to the thermal diffusivity, which in turn is equal to the ratio of the measures of two properties of the medium in which heat flow occurs, namely, the conductivity and the capacity per unit volume. For a system of given geometrical dimensions the diffusivity gives rise to a single time constant, namely, the ratio of the thermal capacitance to the thermal conductance.

The representation described by the Fourier equation can accordingly be expressed, at least in principle, by an equivalent electrical circuit involving only resistances and capacitances. Systems which involve convective as well as conductive flow may be more complicated. Frank-Kamenetskii (1938), for example, claimed that two time constants are associated with thermal convection. The nature of the initial heat flow associated with convection, especially in liquid systems, is formally quite different from that from a resistance-capacity circuit, and one of the authors (Bosworth 1946, 1948) has suggested that these systems exhibit a phenomenon properly called thermal inductance. These phenomena are connected with the temperature changes which follow the sudden steady application of heat to a wire immersed in the fluid. The heat flow from the wire to the fluid was constant, while the temperature difference between the wire and the fluid, which was initially zero, finally settled down to a steady value after a sufficient time of heating. In the phenomenon that one of the authors described as thermal inductance the temperature difference at first rose to a value higher than the final value and approached it from above.

* School of Chemistry, University of New South Wales, Sydney.

† School of Mathematics, University of New South Wales, Sydney.

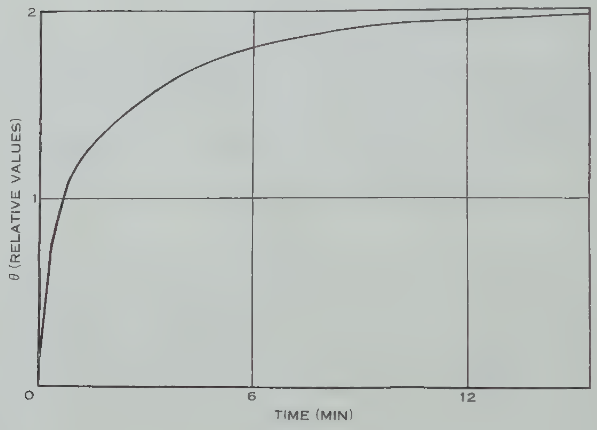

2
1
0
6
12
θ (RELATIVE VALUES)
TIME (MIN)

Fig. 1

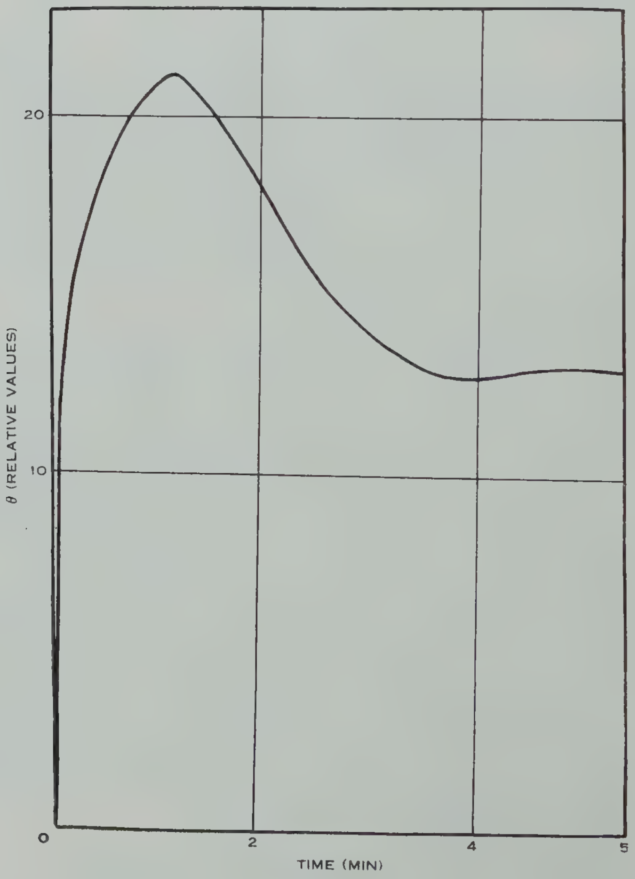

20
10
0
2
4
5
θ (RELATIVE VALUES)
TIME (MIN)

Fig. 2

In certain fluids, notably in gases and in very viscous liquids, the initial rise in the temperature difference above the final value was not realized ; in other words, the final temperature value was approached from below. Such a transient may be regarded as of a purely capacitive type. On the other hand the transient which approaches its full value from above could be described as an inductive type. Thermal transients of both types have been described by Ostroumoff (1956) and interpreted in a purely mechanical manner. A more exhaustive study has shown that moderately viscous liquids appear to give a capacitive-type transient at a low heating rate and an inductive-type transient at a higher heating rate.

Typical transients for each type are shown by a hot wire of 30 S.W.G. immersed in an 81·2 per cent. (supersaturated) sucrose solution at 18·0 °C. Figure 1 shows the thermal transient for a heating current of 0·52 A and Figure 2 shows that for a heating current of 2·65 A. The former appears to be of the capacitive type, the latter of the inductive type.

II. An Assumed Equivalent Circuit

On the application of the flow of heat from the wire to the fluid an initial effect results in setting up a temperature difference between the wire and the outer regions of the fluid, and later also a temperature difference occurs between the fluid in close contact with the wire and the more distant regions. At the same time, there exists initially no circulating system of convection currents and the thermal resistance opposing heat flow is initially high. When the system of convection currents has been fully developed the resistance to thermal flow is considerably reduced. The equivalent circuit representing the heat flow thus must involve two resistances in parallel, one of which r is always operative and the other R only becomes operative after the system of convection currents has been developed. In the equivalent circuit we represent R as in series with an inductance L. The whole system has an additional capacitance C which operates in parallel to r and $R+L$. This takes into account both the heating up processes and the action of setting up a steady temperature gradient through the fluid. The capacitance C is expected to include the thermal capacity of the wire and of the fluid adjacent to it.

The overall suggested equivalent circuit is represented in Figure 3. The circuit, as an electrical one, has been discussed by Carter (1944) in connexion with the transient voltage across the Tirrill regulator contacts on a generator. Carter expressed the input impedance in terms of Heaviside operators and interpreted the operators in terms of exponential functions. For certain ranges of the variables (r, R, C, and L) the exponents become complex quantities and the transient takes the form of a damped oscillation. When the exponents are both real the disturbance is overdamped and there are no oscillations.

By matching actual values of the thermal transient to the theoretical values of the calculated transient with selected values of R, C, L, and r, it is, at least in principle, possible to derive all these associated circuit characteristics. In practice it is relatively easy to match the circuit characteristics from the oscillatory transient. Points may be selected from values of the temperature which pass through the asymptotic value and which attain one or more turning points. However, the derivation of such characteristics as these from the theoretical expressions for the circuit diagram is not particularly convenient in the form given by Carter, and an equivalent, but more convenient, expression may be derived.

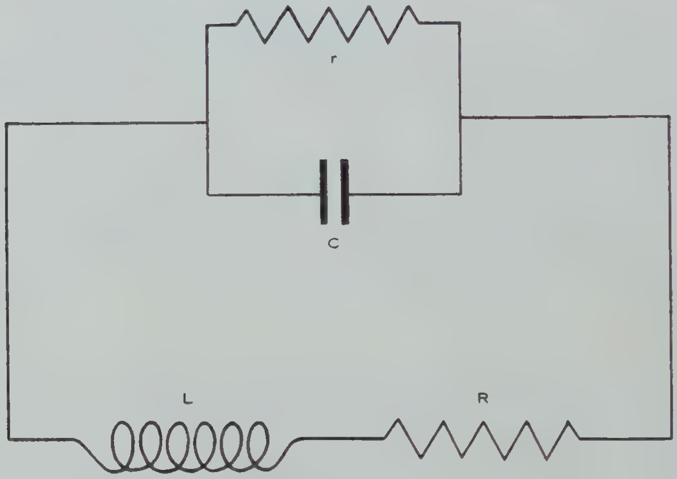


Fig. 3

III. THE CHARACTERISTICS OF THE EQUIVALENT CIRCUIT

The overall formal impedance Z of Figure 3 may readily be seen to be

$$Z=\frac{r(R+Lp)}{CLrp^2+RCrp+Lp+r+R}, \qquad \ldots\ldots\ldots\ldots \quad (1)$$

where p is the Heaviside operator

$$p^{-1}=\int dt.$$

Let the current take the form of the Heaviside function, namely

$$I=0, \text{ when } t<0,$$
$$I=1, \text{ when } t\geqslant 0.$$

The potential for any positive value of t then takes the form ZI, where

$$Z=\frac{rR}{r+R}\left\{1-\exp-\left(\frac{L+RCr}{2rCL}\right)t\left[\cosh\frac{\sqrt{\{(L-RCr)^2-4r^2CL\}}}{2rCL}t\right.\right.$$
$$\left.\left.-\frac{2rL/R+L-RCr}{\sqrt{\{(L-RCr)^2-4r^2CL\}}}\sinh\frac{\sqrt{\{(L-RCr)^2-4r^2CL\}}}{2rCL}t\right]\right\}.$$
$$\ldots\ldots\ldots\ldots \quad (2)$$

This equation may be written in a more concise form by means of the following substitutions:

$$X=(L-RCr)/2r\sqrt{(LC)}, \qquad \text{(2a)}$$

$$Y=(1/R)\sqrt{(L/C)}, \qquad \text{(2b)}$$

and

$$\tau=t/\sqrt{(LC)}, \qquad \text{(2c)}$$

Equation (2) then becomes

$$Z=\frac{rR}{r+R}\left\{1-\exp-\left(X+\frac{1}{Y}\right)\tau\left[\cosh\sqrt{(X^2-1)}\tau-\frac{X+Y}{\sqrt{(X^2-1)}}\sinh\sqrt{(X^2-1)}\tau\right]\right\}. \qquad \text{(3)}$$

When the modulus of X is less than unity the arguments of the hyperbolic function become complex and may be replaced by trigonometrical functions,

$$Z=\frac{rR}{r+R}\left\{1-\exp-\left(X+\frac{1}{Y}\right)\tau\left[\cos\sqrt{(1-X^2)}\tau-\frac{X+Y}{\sqrt{(1-X^2)}}\sin\sqrt{(1-X^2)}\tau\right]\right\}, \qquad \text{(4)}$$

valid when $-1\leqslant X\leqslant 1$.

The oscillatory type of transient is represented by equation (4). A completely overdamped transient is represented by equation (3) with the value X negative and numerically greater than unity. The positive values of X that are greater than unity show no true oscillatory motion. However, the expression in the square brackets in equation (3) can always change from a positive to a negative value with an increasing τ; this means that the transient value of Z overshoots the asymptotic value and finally approaches that value from above and not from below as in the region for which $-1<X<+1$.

Experimentally we may expect a capacitive type of transient to appear when $X<-1$ and an inductive-type transient when $X>-1$. In the limited region $-1\geqslant X\geqslant 1$ the transient may show a series of damped waves. However, in all thermal flow circuits we expect the damping to be fairly severe and any wave other than the first may well be swamped by the noise level.

As indicated earlier the four most readily attainable characteristics of experimental " inductive-type " transients are the value t_0 of the time required to first attain a value equal to the asymptotic value, and the value t_1 to attain the first stationary point and the corresponding impedances Z_0, the asymptotic value, and Z_1, the value at the first maximum. Since four variables are necessary to describe the transient, the four easy experimental values are, at least in principle, sufficient to fix the nature of the transient.

The value of t_1 from equation (3) takes the value

$$\frac{t_1}{\sqrt{(LC)}}=\frac{1}{\sqrt{(X^2-1)}}\,\text{arc tanh}\,\frac{\sqrt{(X^2-1)}}{X} \qquad \text{(5)}$$

$$=\frac{1}{\sqrt{(X^2-1)}}\,\text{arc cosh}\,X. \qquad \text{(5a)}$$

From equation (4) the corresponding value reads

$$\frac{t_1}{\sqrt{(LC)}}=\frac{1}{\sqrt{(1-X^2)}}\text{ arc cos } X. \qquad \dots\dots\dots\dots\dots\dots \text{(6)}$$

The value of t_0 from equation (3) takes the value

$$\frac{t_0}{\sqrt{(LC)}}=\frac{1}{\sqrt{(X^2-1)}}\text{ arc tanh }\frac{\sqrt{(X^2-1)}}{X+Y}, \qquad \dots\dots\dots \text{(7)}$$

and from equation (4)

$$\frac{t_0}{\sqrt{(LC)}}=\frac{1}{\sqrt{(1-X^2)}}\text{ arc tan }\frac{\sqrt{(1-X^2)}}{X+Y}. \qquad \dots\dots\dots \text{(8)}$$

The ratio of these two times, which we shall take as equal to A, becomes, in the trigonometrical region,

$$\frac{t_1}{t_0}=\frac{\text{arc cos } X}{\text{arc tan }\{\sqrt{(1-X^2)}/(X+Y)\}}=A>1.$$

We now make a substitution for X, putting

$$X=\cos T,$$

and obtain

$$Y=\frac{\sin\,(1-1/A)T}{\sin T/A}. \qquad \dots\dots\dots\dots\dots\dots \text{(9)}$$

In the hyperbolic region a corresponding substitution,

$$X=\cosh T,$$

gives

$$Y=\frac{\sinh\,(1-1/A)T}{\sinh T/A}. \qquad \dots\dots\dots\dots\dots \text{(9a)}$$

Let the ratio of the corresponding two impedances be

$$Z_1/Z_0=B.$$

The trigonometrical region then gives

$$B=1+Y\exp-\left\{\frac{X+1/Y}{\sqrt{(1-X^2)}}\text{ arc cos } X\right\}, \qquad \dots\dots \text{(10)}$$

and, in the hyperbolic region,

$$B=1+Y\exp-\left\{\frac{X+1/Y}{\sqrt{(X^2-1)}}\text{ arc cosh } X\right\}. \qquad \dots\dots \text{(10a)}$$

Substituting T for X we get

$$\ln Y/(B-1)=\frac{T}{\sin T}\left(\cos T+\frac{1}{Y}\right), \qquad \dots\dots\dots\dots \text{(11)}$$

or

$$\ln Y/(B-1)=\frac{T}{\sinh T}\left(\cosh T+\frac{1}{Y}\right), \qquad \dots\dots \text{(11a)}$$

in the respective region.

Elimination of Y between equation (9) and (11) now gives

$$\ln (B-1)=\ln \left[\frac{\sin (1-1/A)T}{\sin T/A}\right]-T \cot (1-1/A)T, \quad \ldots. \quad \textbf{(12)}$$

or

$$\ln (B-1)=\ln \left[\frac{\sinh (1-1/A)T}{\sinh T/A}\right]-T \coth (1-1/A)T, \quad .. \quad \textbf{(12a)}$$

in which one of the required combinations of the circuit characteristics (namely, $T=\text{arc}\cos X$ or $\text{arc}\cosh X$) are implicitly expressed in terms of measurable characteristics of the transients, namely, A and B.

IV. EVALUATION OF CIRCUIT PARAMETERS FROM THE CHARACTERISTICS OF THE TRANSIENT

The relationship between X and A and B can be obtained from tables, such as Table 1, derived from equations (**12**) and (**12a**). Reading off from values of A and B (interpolated, if necessary) it is possible to derive X and thus $(L-RCr)/2r\sqrt{(LC)}$ from Table 1.

Table 1 in effect gives a relationship between the parameters A and B at a series of constant values of X. A more convenient tool is a relationship between A and X at a series of fixed values of B. Such data, derived from Table 1, are shown in Figure 4. It will be noted that the curves showing A versus X at constant B are practically rectilinear at the higher values of X. For limiting high values beyond the range covered by Table 1 or Figure 4, equation (**12a**) may be written in the approximate form

$$\ln (B-1)=\ln \frac{e^{(1-1/A)T}}{e^{T/A}}-T,$$

or

$$\begin{aligned} B-1&=e^{-2T/A} \\ &=(2X)^{-2/A}. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad \textbf{(13)} \end{aligned}$$

Under similar conditions the limiting value of Y becomes

$$Y=X^{2(1-2/A)}, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad \textbf{(13a)}$$

so that the apparently linear relationship between A and X does not apply to still higher ranges.

V. THE RATIO OF CONDUCTIVE TO CONVECTIVE RESISTANCE

By using Figure 4 and equations (**9**) (or (**9a**)) it is possible to derive both X and Y from the known values of A and B. Now X and Y are connected from their primary definitions in equations (**2a**) and (**2b**) and, in fact,

$$2X/Y+1/Y^2=R/r. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad \textbf{(14)}$$

Accordingly, the ratio R/r thus depends only on the parameters A and B. The relationship proves to be a particularly simple one. Substituting equation (**9**) for equation (**14**) gives

$$\frac{2 \cos T \sin T/A}{\sin (T-T/A)}+\frac{\sin^2 T/A}{\sin^2 (T-T/A)}=n,$$

TABLE 1

VALUES OF B FOR GIVEN X AND A, COVERING BOTH TRIGONOMETRICAL AND HYPERBOLICAL REGIONS

A X	1·2	1·3	1·4	1·5	1·8	2·0	2·5	3·0	3·5	4·0	4·5
—1·0	1·004+	1·029	1·082	1·163	1·575	2·000	3·776	7·135	13·248	24·141	43·275
—0·9	1·002+	1·015	1·046	1·093	1·324	1·539	2·276	3·297	4·619	6·258	8·237
—0·8	1·002—	1·012	1·037	1·077	1·264	1·435	1·991	2·706	3·567	4·563	5·687
—0·7	1·001+	1·011	1·032	1·066	1·228	1·373	1·831	2·397	3·051	3·779	4·573
—0·6	1·001+	1·0095	1·028	1·059	1·203	1·330	1·726	2·199	2·732	3·312	3·933
—0·5	1·001+	1·009	1·026	1·054	1·184	1·298	1·648	2·059	2·513	3·000	3·513
—0·4	1·001+	1·008	1·024	1·049	1·169	1·273	1·588	1·953	2·351	2·773	3·213
—0·3	1·001—	1·007	1·022	1·046	1·156	1·252	1·540	1·870	2·226	2·600	2·987
—0·2	1·001—	1·007	1·021	1·043	1·146	1·235	1·501	1·803	2·126	2·463	2·811
—0·1	1·001—	1·006	1·0195	1·040	1·137	1·221	1·468	1·747	2·044	2·352	2·669
0	1·001—	1·006	1·018	1·038	1·129	1·208	1·440	1·699	1·975	2·260	2·551
0·1	1·001—	1·006	1·017	1·036	1·122	1·197	1·415	1·659	1·916	2·181	2·452
0·2	1·001—	1·005	1·017	1·034	1·116	1·187	1·393	1·623	1·865	2·114	2·368
0·3	1·001—	1·005	1·016	1·033	1·111	1·178	1·374	1·592	1·821	2·056	2·294
0·4	1·001—	1·005	1·015	1·031	1·106	1·170	1·357	1·564	1·781	2·004	2·231
0·5	1·001—	1·005	1·0145	1·030	1·101	1·163	1·342	1·539	1·746	1·959	2·175
0·6	1·001—	1·005	1·014	1·029	1·097	1·157	1·328	1·517	1·715	1·918	2·124
0·7	1·001—	1·004	1·013	1·028	1·094	1·151	1·315	1·497	1·687	1·882	2·078
0·8	1·001—	1·004	1·013	1·027	1·090	1·145	1·304	1·479	1·661	1·848	2·038
0·9	1·001—	1·004	1·0125	1·026	1·087	1·140	1·293	1·462	1·638	1·817	2·002
1·0	1·0005	1·004	1·012	1·025	1·085	1·135	1·282	1·446	1·616	1·792	1·968
1·2	1·0005—	1·0035	1·011	1·023	1·079	1·127	1·266	1·418	1·579	1·742	1·909
1·4	1·0004+	1·0035	1·011	1·022	1·074	1·120	1·251	1·395	1·546	1·700	1·857
1·6	1·0004+	1·003	1·010	1·021	1·070	1·113	1·237	1·374	1·518	1·664	1·814
1·8	1·0004—	1·003	1·0095	1·020	1·067	1·107	1·225	1·356	1·493	1·633	1·775
2·0	1·0004—	1·003	1·009	1·019	1·064	1·102	1·215	1·339	1·470	1·605	1·741
2·5	1·0003+	1·0026	1·0081	1·017	1·057	1·091	1·193	1·305	1·424	1·547	1·672
3·0	1·0003+	1·0024	1·0073	1·015+	1·051	1·083	1·175	1·279	1·388	1·502	1·617
3·5	1·0003—	1·0022	1·0067	1·014	1·047	1·076	1·161	1·257	1·359	1·465	1·573
4·0	1·0003—	1·0020	1·0061	1·012+	1·043	1·070	1·149	1·239	1·335	1·435	1·537
4·5	1·0002+	1·0019	1·0057	1·012—	1·040	1·065	1·139	1·224	1·314	1·409	1·506
5·0	1·0002+	1·0017	1·0053	1·011—	1·037	1·060	1·130	1·211	1·297	1·387	1·480
5·5	1·0002+	1·0016	1·0049	1·010+	1·035	1·057	1·123	1·199	1·282	1·368	1·458
6·0	1·0002—	1·0015	1·0046	1·010—	1·033	1·053	1·116	1·189	1·268	1·352	1·438
6·5	1·0002—	1·0014	1·0043	1·009—	1·031	1·050	1·110	1·180	1·256	1·337	1·420
7·0	1·0002—	1·0014	1·0041	1·008+	1·029	1·048	1·105	1·172	1·246	1·324	1·405
7·5	1·0002—	1·0013	1·0039	1·008+	1·028	1·045	1·100	1·165	1·236	1·312	1·390
8·0	1·0002—	1·0012	1·0037	1·008—	1·026	1·043	1·096	1·159	1·228	1·301	1·377
8·5	1·0001+	1·0012	1·0035	1·007+	1·025	1·041	1·092	1·153	1·220	1·291	1·366
9·0	1·0001+	1·0011	1·0034	1·007—	1·024	1·040	1·089	1·147	1·213	1·282	1·355
9·5	1·0001+	1·0011	1·0032	1·007—	1·023	1·038	1·085	1·142	1·206	1·274	1·345
10·0	1·0001+	1·0010	1·0031	1·006	1·022	1·037	1·082	1·138	1·200	1·267	1·336

with n the ratio of the resistances. This expression may readily be reduced to

$$\sin^2 T=(1-n)\sin^2 T(1-1/A), \qquad \ldots\ldots\ldots\ldots \quad (\mathbf{15})$$

with a corresponding expression in the hyperbolic region.

The quantity n may take values from 0 to ∞ but may never be negative. The upper limit corresponds to a type of heat flow in which the contribution by the convective transfer has vanished. The lower limit corresponds to a condition at which the convective contribution is infinitely large in comparison with the conductive transfer. The only solution of equation (**15**) pertinent to practical problems for which n vanishes is the particularly simple form

$$A=T/(2T-\pi), \qquad \tfrac{1}{2}\pi<T<\pi. \qquad \ldots\ldots\ldots\ldots \quad (\mathbf{16})$$

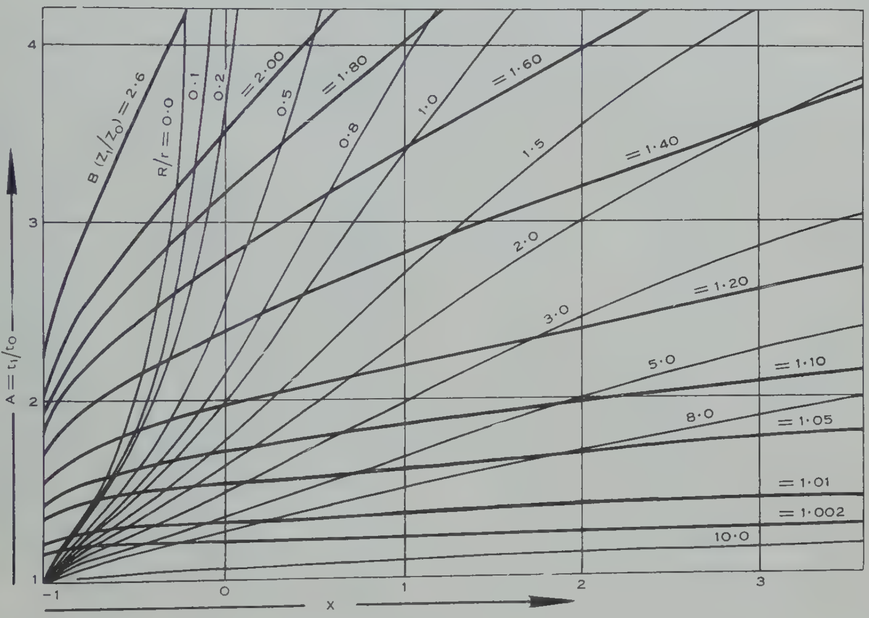


Fig. 4

The points where the curve of $R/r=0$ on Figure 4 cross the lines of constant B are solutions of equation (**16**). The region to the right of this line corresponds to finite values of R/r, that to the left to unattainable values of R/r. Any experimental value of A and B lying to the left of this line thus corresponds to a condition which is inconsistent with the equivalent circuit postulated in Figure 3.

Values for A at a range of values of T (or X) may be obtained from equation (**15**) for selected values of n (0·1, 0·2, etc.). The lines at selected values of n are drawn as a series of contours crossing the constant B lines in Figure 4. Accordingly, from this figure we can read off, given a known pair of values for A and B, the corresponding values for X and for the ratio R/r.

VI. TREATMENT OF EXPERIMENTAL RESULTS

Experimental results are obtained in the form of a transient curve plotting the temperature excess (or alternatively the quotient of the temperature excess by the heat flow rate) against the time. The peak temperature and the time to attain that peak are recorded, together with the final temperature and the time first to cross that amplitude. From these measurements the ratio of temperatures (B) and the ratios of the times (A) can be read off. From Figure 4 we may now read off the value X and also the ratio R/r. Alternatively, from the X value we may compute the value of Y using equation (**9**) or (**9a**)). From Y and X the value of R/r can also then be derived by use of equation (**14**). From the final temperature

$$Z_0=Rr/(R+r), \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{17})$$

it is then possible to obtain values for both R and r.

TABLE 2

VALUES OF THE RATIO $\sqrt{(LC)}/t_1$ AS FUNCTION OF X

X	$\sqrt{(LC)}/t_1$	X	$\sqrt{(LC)}/t_1$	X	$\sqrt{(LC)}/t_1$
—1·0	0·0000	0·6	0·8627	5·5	2·263
—0·9	0·1620	0·7	0·8978	6·0	2·387
—0·8	0·2402	0·8	0·9324	6·5	2·510
—0·7	0·3044	0·9	0·9665	7·0	2·630
—0·6	0·3613	1·0	1·0000	7·5	2·749
—0·5	0·4135	1·2	1·066	8·0	2·866
—0·4	0·4624	1·4	1·130	8·5	2·983
—0·3	0·5086	1·6	1·193	9·0	3·098
—0·2	0·5529	1·8	1·255	9·5	3·211
—0·1	0·5954	2·0	1·315	10·0	3·324
0	0·6366	2·5	1·462	12·0	3·763
0·1	0·6766	3·0	1·605	14·0	4·191
0·2	0·7155	3·5	1·743	16·0	4·608
0·3	0·7534	4·0	1·877	18·0	5·018
0·4	0·7906	4·5	2·008	20·0	5·415
0·5	0·8270	5·0	2·137		

Again from equation (**2b**) it is possible to obtain a direct measurement of $\sqrt{(L/C)}$,

$$\sqrt{(L/C)}=RY.$$

A complementary expression may be obtained for the quantity $\sqrt{(LC)}$. The ratio $t_1/\sqrt{(LC)}$ using equations (**5a**) and (**6**), may be expressed as a function of X. Numerical data for this function (covering the range $-1\leqslant X<10$) are given in Table 2.

From this table the values of $\sqrt{(LC)}$ may be derived from t, knowing the value of X. With a very high value of X the expression takes a limiting form

$$\sqrt{(LC)}=Xt_1/2\cdot303 \log 2X. \qquad \ldots\ldots\ldots\ldots \quad (\mathbf{18})$$

From known values of $\sqrt{(L/C)}$ and $\sqrt{(LC)}$, independent values of L and C may be obtained, and the identification of all experimental parameters is now complete. Examples illustrating these methods will be given in the following paper (Bosworth 1960).

VII. REFERENCES

BOSWORTH, R. C. L. (1946).—*Nature* **158** : 309.
BOSWORTH, R. C. L. (1948).—*Nature* **161** : 166.
BOSWORTH, R. C. L. (1960).—*Aust. J. Phys.* **13** : 84.
CARTER, G. W. (1944).—" The Simple Calculation of Electrical Transients." (Cambridge Univ. Press.)
FRANK-KAMENETSKII, D. A. (1938).—*C.R. Acad. Sci. U.R.S.S.* **18** : 409.
OSTROUMOFF, C. A. (1956).—*J. Tech. Phys., Moscow* **26** : 2720.

ATTEMPTS TO MEASURE THE INDUCTIVE ELEMENT ASSOCIATED WITH THE NATURAL CONVECTION OF HEAT

By R. C. L. Bosworth*

[*Manuscript received November* 6, 1959]

Summary

A study has been made of the variation in time of the temperature of a wire immersed in a fluid and heated by a constant electric current. For a given fluid the curve obtained by plotting the ratio of the temperature of the wire to the heat input versus the time is initially the same shape for all rates of heat input. Divergences from the lowest heating rate set in only when the system of convection currents sets in. This occurs at earlier times after the commencement of heating the higher the heating rate. Expressions already developed are used to evaluate the resistive, capacitive, and inductive elements required to fit the observed transient curves. The values of the former two types of element are consistent with an assumed stagnant film of a thickness the order of 1 mm around the heated wire, but the value of the deduced inductive element is some 10^4–10^5 greater than that associated with the kinetic energy belonging to the system of convection currents.

I. Introduction

The work described in the previous paper (Bosworth and Groden 1960) has provided a means for evaluating certain thermal properties of a fluid from the study of convective heat transfer in the transient state. The object of the present paper is to describe some preliminary exploration of the use of the methods derived, to show the broad effects of experimental conditions, to evaluate the accuracy of the derived quantities, and to find the optimum conditions for accuracy in further work.

As a result of a complete analysis of the transient curve, four parameters associated with thermal transfer from a hot solid to the surrounding liquid could be obtained. These include two thermal resistance terms, one associated with a purely conductive flow (r in the paper cited) and one associated with a shunt flow due to convection alone (R). The other two parameters derived are C, a capacitance term and probably a measure of the thermal capacity of the stagnant film around the heated solid, and L, an inductance term tentatively associated with the kinetic energy stored in the system of circulating fluid around the heated body.

II. Experimental Details

The system selected for experimental study consisted of a heated resistance alloy wire which was held horizontally by stout current lead and immersed in the fluid under test. The wire was heated by an A.C. current from a source of steady e.m.f. through a variable resistance and current meter. The wires used

---

* School of Chemistry, University of New South Wales, Sydney.

varied in diameter from 0·610 to 0·274 mm (23–32 S.W.G.) and the lengths varied from 1·5 to 2·5 cm. The heat flux density for unit length of a wire heated by a current I of diameter d and specific (electrical) resistance ρ is given by

$$q=4\rho I^2/J\pi^2d^3, \qquad (1)$$

with J the Joule's equivalent. For a fixed experimental set-up q is thus directly proportional to I^2.

The temperature difference between the wire and the fluid (θ) was obtained by means of a thermocouple of wire at least 3 gauges finer than the heater wire. One junction of this couple was welded to the centre of the heater wire and the other (immersed in the fluid) was situated some 5 mm away and in the same horizontal plane. The thermocouple was connected through a variable shunt box to a recording galvanometer. Relative measures of θ can be obtained from the galvanometer reading and absolute readings can be obtained from calibration of a like couple in baths of different known temperatures.

The liquids examined were water, aqueous solutions, and various organic liquids. Some 200 ml was used in any test. No special purification was attempted. The vessel containing the liquid under test and the hot wire thermocouple device were contained in an external thermostat at a temperature ranging from 12 to 70 °C and the vessel was allowed to come to thermal equilibrium before any measure was attempted. A predetermined heating current was then switched on and the temperature difference recorder was allowed to run until the temperature became steady. When the time taken was in excess of 10 min an ordinary bench reflecting galvanometer with frequent reading could be used. Processes which could have been completed in less than 5 sec have not yet been examined, although there appears to be evidence that such fast transients could be obtained under certain conditions. Particularly quick acting recorders are necessary in the analysis of these transients and these have not yet been used in this field.

The experimental factors which might be expected to affect the nature of the convective transient are the diameter of the wire, its temperature excess θ, and the thermal nature of the fluid, including the thermal conductivity and the convective modulus. Changes in θ provide the easiest experimental variation, limitation being set only by the onset of boiling at the upper limit and the sensitivity of the galvanometer at the lower limit. However, as will be seen later, there is a great tendency for the time constants to become excessively long at low values of θ and the ultimate limiting factors may be the stability of the heating current and of the thermostat bath. Changes in d with the present form of device are limited to a comparatively narrow range. Below 32 S.W.G. it is difficult to get suitable thermocouples which do not produce an appreciable thermal disturbance at the centre of the heated wire. Wires greater than 22 S.W.G. tend to draw an excessive heating current and seriously disturb external thermostating. The range of thermal conductivity in non-metallic liquids is comparatively restricted (Bosworth 1952). On the other hand there are readily available fluids covering a wide range of convective moduli: this modulus is an experimental variable which may be readily changed by using viscous liquids at different compositions or temperatures. Mixtures used include the systems glycerol–water and sucrose–water.

III. Evaluation of Experimental Data

As already intimated, relative values of the heat flux q and the temperature difference θ may be obtained more readily than absolute figures. However, the first step in the deduction of the circuit parameters is concerned solely with dimensionless quantities, A and B, which are ratios of times and impedances respectively as derived from the transient curves. From A and B, using data from the tables and curves given in the previous paper (Bosworth and Groden 1960), it is possible to derive the quantities X, Y, and R/r, all of which are dimensionless. To proceed further it is necessary to take into account the absolute value of t_1 and then get the quantity $\sqrt{(LC)}$, or alternatively to take into account the final thermal impedance Z_0 which is equal to $Rr/(R+r)$ or θ/I^2 for any fixed experimental set up. From the known absolute (or relative) values of θ/q it is then possible to derive R and r. From

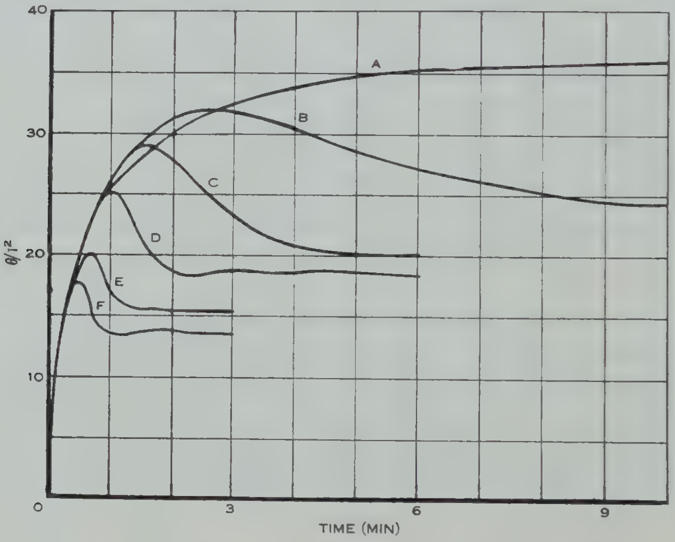


Fig. 1.—Transient curves, plotted as θ/I^2 against time.

$$\sqrt{(L/C)}=YR, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

and from $\sqrt{(LC)}$ it is then possible to evaluate both L and $1/C$ on the same (or relative) scale.

As an illustration we will consider data derived from a wire of diameter 0·05105 cm immersed in 60·4 per cent. w/w sucrose solution at 12 °C. The heating current ranged from 0·230 A through 0·362, 0·540, 0·790, 1·220, to 1·590 A. The transient curves, plotted in the form of θ/I^2 versus time, are shown in Figure 1. The lowest heating rate gave a curve with a very indistinct maximum, at least within the time of the experiment recorded, while the curves associated with the higher heating rates showed transients passing successively faster and faster through a maximum and then approaching the final value from above. Some of the curves show signs of a severely damped oscillation after passing through the first maximum.

From readings taken from the third curve we obtain the following four characteristics :

$$\left.\begin{array}{ll} t_0 & 27 \text{ sec} \\ t_1 & 92 \text{ sec} \end{array}\right\} \quad A=3\cdot 40,$$

$$\left.\begin{array}{ll} Z_0 & 20\cdot 4 \text{ arbitrary units} \\ Z_1 & 29\cdot 9 \text{ arbitrary units} \end{array}\right\} \quad B=1\cdot 415.$$

From Figure 4 of the previous paper it is now possible to read off both

$$X=2\cdot 3 \text{ and } R/r=1\cdot 80,$$

from which

$$Y=2\cdot 75.$$

From the value of Z_0 (measured in arbitrary units as θ/I^2) we then arrive at the values

$$R=57\cdot 2 \text{ and } r=31\cdot 7 \text{ arbitrary units.}$$

From the values of R and Y we obtain

$$\sqrt{(L/C)}=157 \text{ in the same arbitrary units,}$$

and finally using the value of t_1 and the ratio $\sqrt{(LC)}/t_1$, depending on X and reading off from Table 2 of the previous paper, we get

$$\sqrt{(LC)}=131 \text{ sec.}$$

Accordingly

$$L=20600 \text{ arbitrary units, and}$$
$$C=0\cdot 83 \text{ reciprocal units.}$$

IV. Experimental Results

Analysis of the curves shown in Figure 1 may be carried out by the method outlined above. However, it is more convenient to convert the purely relative figures to absolute figures from the known diameter and electrical resistivity of the wire concerned, using equation (**1**). For a wire of diameter 0·05105 cm with specific resistivity $\rho=10^{-4}\,\Omega$ cm, the heat flux per unit length becomes

$$q=0\cdot 071 I^2 \text{ cal cm}^{-1}\text{ sec}^{-1}, \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{3})$$

so that in turn $rR/(R+r)$ becomes

$$(\theta/q)_{\text{final}}=14(\theta/I^2)_{\text{final}}.$$

The appropriate values of the circuit parameters derived from the six curves shown in Figure 1 are given in Table 1, while Table 2 contains corresponding values from the same wire in different solutions and Table 3 for different wires in various liquids.

TABLE 1

THERMAL CIRCUIT PARAMETERS ASSOCIATED WITH A WIRE OF DIAMETER 0·0511 CM IN 60·4 PER CENT. SUCROSE SOLUTION AT 12 °C

θ (degC)	r (degC cm sec cal^{-1})	R (degC cm sec cal^{-1})	L (degC cm sec^{-2} cal^{-1})	C (cal (degC)$^{-1}$ cm^{-1})
1·7	540	2700	25,000,000	0·160
3·15	465	1200	1,300,000	0·065
5·9	445	800	280,000	0·060
11·5	400	750	160,000	0·049
23·2	376	535	35,000	0·041
34·4	364	410	12,000	0·054

TABLE 2

THERMAL CIRCUIT PARAMETERS ASSOCIATED WITH A WIRE OF 0·0511 CM DIAMETER IN SUCROSE SOLUTION

Strength of Solution	Bath Temperature	θ (degC)	r (degC cm sec cal^{-1})	R (degC cm sec cal^{-1})	L (degC cm sec^2 cal^{-1})	C (cal (degC)$^{-1}$ cm^{-1})
60·4%	16 °C	2·6	440	2300	1,700,000	0·072
		3·7	380	1700	860,000	0·049
		5·5	400	1260	410,000	0·045
		9·2	460	1200	170,000	0·044
		12·8	440	780	78,000	0·049
		18·5	410	700	53,000	0·036
		25·5	390	700	39,000	0·030
		34	390	490	11,000	0·045
		59	380	350	3,100	0·046
		63	440	280	2,500	0·035
		79	520	170	710	0·037
		86	620	63	510	0·048
		112	320	280	1,400	0·029
		138	245	190	1,100	0·046
	40·5 °C	3·4	650	2600	2,740,000	0·048
		7·6	460	590	49,000	0·043
		13·0	320	680	19,000	0·043
		42	290	360	2,900	0·041
	69 °C	5·6	420	1400	240,000	0·022
		17·1	500	400	24,000	0·022
		24·6	430	300	2,000	0·037
		45·6	490	200	410	0·021
81·2%	18 °C	132	520	520	12,800	0·050
55·4%	19 °C	6·5	430	640	4,200	0·045
		40·0	340	250	220	0·054

TABLE 3

THERMAL CIRCUIT PARAMETERS FROM VARIOUS SOURCES

Wire Diameter (mm)	Liquid	Bath Temperature	θ (degC)	r (degC cm sec cal^{-1})	R (degC cm sec cal^{-1})	L (degC cm sec^2 cal^{-1})	C (cal (degC)$^{-1}$ cm^{-1})
0·457	40% sucrose solution*	19 °C	2·8 4·3	530 840	680 400	53,000 8,700	0·083 0·077
0·457	70% sucrose*	—	47	440	1160	285,000	0·044
0·1016	water	22 °C	26	12·1	42·1	2,700	0·168
0·1016	carbon tetrachloride (?)	—	4·0	90	330	280,000	0·084
0·1016	toluene	—	2·4 3·1 4·4 37	64 61 65 55	620 510 350 130	18,000 3,800 3,500 2,600	0·170 0·089 0·092 0·090

* Data from published curves (Bosworth 1946*a*).

V. ACCURACY

Before considering possible relationships between the circuit parameters and the properties of the heated wire and the convective medium, some thought must be given to possible sources of error. Two obvious sources of inaccuracy immediately come to mind. The device used cannot strictly be described by any simple mathematical expression. The small thermocouple has been used to measure the temperature difference between the central portion of the hot wire and the surrounding fluid. In practice, of course, the presence of the couple alters the electrical conductivity of the central portion of the wire and so disturbs the rate of heat generation. Again, the arms of the couple carry heat away from the heated zone and so tend again to depress the temperature of the zone for which a temperature measurement is required. Disturbing influences may be reduced by using finer wires for the thermocouple and may become insignificant with sufficiently fine wire. However, in an initial broad study of the allied phenomena this particular aspect has not yet received quantitative attention.

For any particular experimental set-up, assuming that the actual temperature is a representative central temperature of the wire, then the accuracy for all four parameters is largely dominated by the relationship between the A's, or ratio of the times for attaining standard points on the transient, and the B's, or ratio of the temperature differences. Figure 4 in the previous paper shows this relationship. At a value of B equal to 1·2, which is the average value of this experimental ratio recorded in the data used in Table 1, a 1 per cent. error in the ratio A will give rise to 3 per cent. error in the ratio of R/r and thus subsequently in the derived parameters. Again, at a value of A equal to 2·5 (the average of the experimental figures used above) an error of 1 per cent. in the value of B will give rise to a 10 per cent. or larger error in the ratio of R/r and thus in subsequently derived parameters. Accordingly, it is vitally important, if reliable circuit parameters are to be obtained, that the measurements of B should be made with the utmost care. It is indeed fortunate that the amplitudes of transient records can be recorded and measured with considerable accuracy—with a higher accuracy for example, than that which may be recorded for the time required to reach a turning point. In the latter case the flat nature of the maximum and the asymmetrical shape of the peak considerably limits the accuracy of assessment of the true turning point. However, assuming that we can measure Z_1 (or the amplitude at the peak) with a reasonably high accuracy the value of all subsequent data will primarily be limited by the accuracy with which we may measure Z_0 (or the final impedance). For this reason it is most desirable that current stabilizers should be used in the heating circuit. Any fluctuation in the rate of heating will vitiate attempts to attain a highly accurate measurement of B as Z_1/Z_0. Again, if the transient involves a particularly long time the experimentalist may read an apparent value of Z which has in fact not settled down to the final Z_0. Again, if the period is long and as a considerable source of internal heat must be applied to the fluid during the course of the transient, the apparent Z_0 may in fact not be the true Z_0 because the properties of the fluid have been changed. Long-period transients may therefore be expected to give false readings of the thermal circuit parameters.

On the other hand, if the thermal transient is associated with too short a period the galvanometer or recorder may not follow truly the whole course of the transient, particularly as the low resistance associated with the thermocouple circuit may cause serious overdamping of the galvanometer. Errors due to this cause usually first show up in abnormally high values of t_0 (the time taken first to cross Z_0) relative to the value of t_1. Overdamping of the galvanometer gives a spuriously low value of A, which has shown up in many instances in which a bench galvanometer has been replaced by a quick-acting potentiometric recorder which gives an apparently much higher value of A for the same conditions as those studied with a reflecting galvanometer. Generally speaking therefore the method described in these two papers can be applied to very quick-acting transients only if suitable quick-acting recorders are used.

A study of the effect of errors due to mis-estimation of the final value of Z_0 may be carried out on curve A of Figure 1. The final value of Z obviously is attained outside the range of time recorded in the figure given (that is, up to 10 min) and has to be taken as 32 arbitrary units. Had we postulated that at the time of 30 min Z had in fact not yet reached its final value and that instead a longer time would have given $Z_0=31$, then the values of X and Y which had been computed as 11·6 and 4·65 would instead be 14·2 and 10·3, with the associated circuit parameters in the order given in Table 1 : 580, 1620, 42×10^6, and 0·150. Conversely, an assumption that due to heating of the fluid the value of 32 had already overshot the final true value and $Z_0=33$ is a better representation, then the four parameters would have been estimated as at 510, 5050, 14×10^6, and $0\cdot185$.

The apparent abnormally high values of C, R, and L at the lowest heating rate cannot therefore be accepted at this stage and a more carefully designed experiment appears to be required.

VI. Conclusions from Experimental Data

The variation in the four circuit parameters, as shown in Tables 1–3, while subject to fairly large errors, particularly at the lowest heating rate, are sufficiently pronounced to make a series of broad generalizations.

All the four properties appear to change somewhat with change in the nature of the convective liquid and with the size of the heating surface. However, the most pronounced change is that associated with the change in the overall temperature difference θ. The two properties r and C, those which have tentatively been associated with the conductive contribution to the total heat flux, are practically unchanged with change of θ. On the other hand, the properties R and L, tentatively attributed to the convective contribution to the total heat flux, both decrease rapidly with increase of θ, the change in L being particularly pronounced. This behaviour, of course, is, at least in its general characteristics, in accordance with the well-known features of convective transfer. However, a quantitative examination of the magnitude of the quantities concerned could add considerable knowledge to this subject.

The capacity term, for example, appears to vary with the diameter of the wire concerned. An average figure of $0\cdot050$ cal cm^{-1} (degC)$^{-1}$ was obtained for a wire of $0\cdot511$ mm diameter. This figure may be compared with the thermal capacity per centimetre of the wire itself. The corresponding capacity C^1 of the wire is

$$C^1=\tfrac{1}{4}\pi\rho c d^2 \text{ cal cm}^{-1}\text{ (degC)}^{-1},$$

where ρ is the density of the wire, c its specific heat, and d its diameter.

Inserting numerical values we get

$$C^1=1\cdot7\times10^{-3}\text{ cal cm}^{-1}\text{ (degC)}^{-1},$$

which is only about 1/30 of the measured value (C) of the capacity per unit length of the wire. The thermal capacity of whatever is contributing to the effective C as deduced from the nature of the thermal transient is thus much larger than the thermal capacity of the heated solid alone. The thickness of the fluid layer surrounding the wire required to give a capacity of $0\cdot05$ cal cm^{-1} (degC)$^{-1}$, using the Janovsky and Archangelsky (1928) data for the volumetric specific heat of sucrose solution, amounts to $1\cdot1$ mm. Taking the capacity of the toluene films surrounding a wire of $0\cdot1016$ mm diameter as $0\cdot09$ cal cm^{-1} (degC)$^{-1}$ (Table 3) gives a corresponding sheet thickness of $1\cdot0$ mm. These figures may be compared with the estimated film thickness of 1 mm in mild thermal convection given by Eckert and Soehngen (1952).

The value of the parameter r in Tables 1–3 must be associated with the flow of heat from the wire before a condition of convection sets in. This involves an unsteady state problem in heat flow in a cylindrical system and is best not exhaustively treated until more complete and more accurate data have been compiled. Suffice it to say, however, that the solution will depend only on the diameter of the wire (possibly with correction due to the presence of the thermocouple) and on the thermal conductivity of the liquids. Analysis thus offers a possibility of eliminating the disturbing effects of convection on the measurements of conduction in mobile liquids, and this technique may well lead to a quick laboratory method for comparing the thermal conductivities of mobile liquids.

The factors R, which are highly sensitive to the temperature excess θ, are in turn indices to the convective moduli. As such these factors describe the ways in which the additional heat flow rate varies with the overall temperature difference in natural thermal convection.

The factors L were, in the first instance, thought to be interpreted as the measure of the kinetic energy stored in the system of circulating convection currents per unit of squared thermal flux. One is tempted to write, on analogy with the corresponding expression in electrodynamics, for the kinetic energy K,

$$K=\tfrac{1}{2}Lq^2. \qquad \dots\dots\dots\dots\dots\dots\dots \quad (4)$$

This, however, on the basis of the definitions and units so far used, is dimensionally incorrect in just the same way as would the expression

$$V=\tfrac{1}{2}\left(\int q\mathrm{d}t\right)^2/C$$

be dimensionally incorrect for the thermal (potential) energy stored in the capacity.

These formal difficulties could, according to the author (Bosworth 1946*b*), be neatly avoided by a redefinition of the thermal flux as a definition of a flow of entropy. Equation (3) then becomes

$$q=\frac{0\cdot071}{T}I^2 \text{ E.U. cm}^{-1}\text{ sec}^{-1}, \qquad \dots\dots\dots\dots \quad (5)$$

with T the absolute temperature of the sink. On the basis of the units previously used, namely, joules $(\text{degC})^{-1}$ cm^{-1}, q becomes

$$q=\frac{0\cdot30}{T}I^2 \text{ thermal amperes/cm}. \qquad \dots\dots\dots\dots \quad (6)$$

On this basis the various properties listed in Tables 1–3 could then be changed to new units as thermal ohms.cm (or $(\text{degC})^2$ sec cm joules^{-1}) for the second and third columns, thermal henry cm (or $(\text{degC})^2$ sec^2 cm joules^{-1}) for the fourth column, and thermal farads per cm (or joules cm^{-1} $(\text{degC})^{-2}$) for the last. The second, third, and fourth columns should be multiplied by $T/4\cdot2$ and the fifth by $4\cdot2/T$.

Equation (4) now becomes dimensionally correct and K is measured in joules per cm of the heated wire. As we are concerned with the convective contribution to the heat current we put

$$q=\theta/R$$

or

$$K=\tfrac{1}{2}L\theta^2/R^2. \qquad (7)$$

To find the way in which the kinetic energy changes with the temperature difference θ, it is therefore necessary to find the relationship between L/R^2 and θ. Data from the Tables 1 and 2 have been plotted in Figure 2 in the form $\log L/R^2$ versus $\log \theta$. The differently marked points corresponding to different bath temperatures fall on different parallel straight lines, all having the slope $-1\cdot0$. We accordingly put

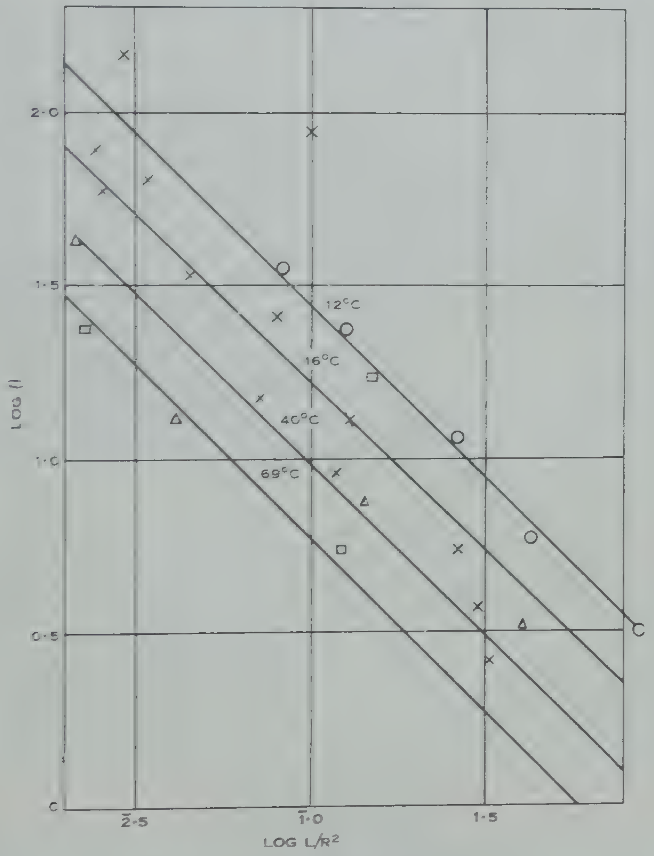


Fig. 2.—Relation between L/R^2 and θ.

$$L/R^2=a\theta^{-1},$$

with the factor a as a constant for any particular wire in any particular fluid. Thus,

$$K=\tfrac{1}{2}a\theta, \qquad (8)$$

so that it appears that the kinetic energy of the convective current increases linearly with the temperature excess. For the 0·511 mm wire in 60·4 per cent. sucrose solution at 12 °C the quantity a takes the value 0·041 J $(\text{degC})^{-1}$ cm^{-1}, and at 16 °C, 0·025 J $(\text{degC})^{-1}$ cm^{-1}. Accordingly, $K = 0\cdot020$ J/cm at 12 °C, and 0·012 J/cm at 16 °C, for a temperature rise of 1 degC and with somewhat lower figures in the less viscous fluids. Now the total volume of fluid in which the hot wire was immersed was some 200 ml, and at the most some 30 ml/cm of wire could be involved in the system of convection currents. The expected kinetic energy in such a system could only be of the order of ergs/cm. The phenomenon giving rise to the inductive features of thermal convection is of one or two orders of magnitude larger than that expected from the kinetic energy of the system of convection chimneys.

Presumably any process in which the heat flow rate is partially conditioned by states of the system arising from absorption of energy in a non-thermal form from the initial heat flow will add an inductive-like element to the thermal circuit. The storage of heat in a process of surface evaporation would act in this way. In some instances cited in Table 2 the liquid was actually caused to boil on the wire (although the liquid medium as a whole did not boil). While in these instances the " inductance " was larger than the extrapolated value from the non-boiling region, however, the change was not one of a different order of magnitude, as might have been expected. Accordingly, there arises the suggestion that the unexpectedly large values associated with thermal induction without surface boiling may be due to surface chemical phase change, for example, to deaeration of the liquid near the wire and subsequent re-solution of microscopic bubbles in the slightly cooler region. This suggests that thermal transients could profitably be studied after complete deaeration of fluids under test.

As an alternative interpretation of the high value of thermal inductance there may be cited our general lack of complete knowledge of the mechanism involved in the passage of layers of different fluid density over a solid surface—the so-called tea-pot effect (Reiner 1956). The tendency of flowing streams of different density to adhere to solid surfaces appears to be several orders of magnitude greater than that expected from normal surface (tension) phenomena. Reiner has suggested that the creation of micro-eddies is responsible. If upheld, a similar mechanism may also be the cause of the apparently high kinetic energy term connected with the act of setting up the convection chimney.

VII. REFERENCES

BOSWORTH, R. C. L. (1946*a*).—*Nature* **163** : 302.

BOSWORTH, R. C. L. (1946*b*).—*Phil. Mag.* **37** : 803–8.

BOSWORTH, R. C. L. (1952).—" Heat Transfer Phenomena." (Assoc. Gen. Publications: Sydney.)

BOSWORTH, R. C. L., and GRODEN, C. M. (1960).—*Aust. J. Phys.* **13** : 73.

ECKERT, E. R. G., and SOEHNGEN, E. (1952).—*Trans. Amer. Soc. Mech. Engrs.* **126** : 227–73.

JANOVSKY, V. V., and ARCHANGELSKY, P. A. (1928).—*Zh. Sakharn. Prom.* **2** : 614–8.

REINER, M. (1956).—*Physics Today* **9** (9) : 16–20.

SHORT COMMUNICATIONS

AN ANALOGY BETWEEN THE EFFECTS OF FLUCTUATING ELECTRIC FIELDS AND STEADY MAGNETIC FIELDS IN ISOTROPIC CONDUCTORS WHEN A UNIVERSAL RELAXATION TIME CANNOT BE DEFINED*

By E. J. MOORE†

It is well known that when a " universal " time of relaxation (τ) exists, the influence of a harmonically varying electric field ($\mathbf{F} \propto e^{i\omega t}$) on the transport properties of a solid may be taken into account by replacing τ by $\tau/(1+i\omega\tau)$. Dingle (1956*a*) demonstrated that, for an isotropic solid, the effect of a steady magnetic field may similarly be obtained by replacing τ by $\tau/(1+j\Omega\tau)$ with an applied d.c. electric field, and by $\tau/[1+(i\omega+j\Omega)\tau]$ with an a.c. field. (Here $j^2=-1$, $ij\neq-1$, and $\Omega=(-\varepsilon)H/mc$ is the circular frequency of precession of an electron.) The object of the present note is to show that this analogy between a high frequency electric field and a steady magnetic field still exists, even when a " universal " relaxation time cannot be defined.

For ease of exposition, we consider a one-band model and assume that the applied fields are an electric field $\mathbf{F}=(F_x,F_y,0)\propto e^{i\omega t}$ and a constant magnetic field $\mathbf{H}=(0,0,H)$. The acceleration of a charge carrier (charge $-\varepsilon$, effective mass m, energy $E=\frac{1}{2}(\hbar/m)^2\,|\,\mathbf{k}\,|^2$, wave vector $\mathbf{k}$, and velocity $\mathbf{v}=\hbar\mathbf{k}/m$) is then

$$\dot{\mathbf{v}}=-\frac{\varepsilon}{m}\left(\mathbf{F}+\frac{1}{c}\mathbf{v}\wedge\mathbf{H}\right).$$

As usual, we start with the Boltzmann equation

$$\frac{\partial f}{\partial t}-\frac{\varepsilon}{m}\left(\mathbf{F}+\frac{1}{c}\mathbf{v}\wedge\mathbf{H}\right)\cdot\nabla_{\mathbf{v}}f=\left[\frac{\partial f}{\partial t}\right]_{\text{coll.}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

and assume a first-order solution of the form (Wilson 1953, p. 210)

$$f=f_0-\mathbf{k}\cdot\mathbf{c}(E)\frac{\partial f_0}{\partial E}. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

Here, f is the electron distribution function, $[\partial f/\partial t]_{\text{coll.}}$ is the rate of change of f due to the various scattering mechanisms, f_0 is the equilibrium distribution function, and $\mathbf{c}(E)$ is a function of the electron energy and the applied fields.

Now, for the more important of those scattering mechanisms for which a universal relaxation time does not exist—acoustic phonon scattering in metals below the Debye temperature (Wilson 1953, p. 263) and optical phonon scattering

---

* Manuscript received December 2, 1959.

† Physics Department, University of Western Australia, Nedlands, W.A.; present address: Physics Department, Harvard University, Cambridge, Mass., U.S.A.

in polar semiconductors (Howarth and Sondheimer 1953)—the collision term may be written

$$\left[\frac{\partial f}{\partial t}\right]_{\text{coll.}} = \mathbf{k}\cdot\mathrm{L}(\mathbf{c})\frac{\partial f_0}{\partial E}, \quad \ldots\ldots\ldots\ldots\ldots \quad (3)$$

where $\mathbf{c}=(c_1,c_2,0)$, $\mathrm{L}(\mathbf{c})=(\mathrm{L}[c_1],\mathrm{L}[c_2],0)$, and L is a linear operator, i.e. if α and β are any functions independent of the electron energy, then

$$\mathrm{L}(\alpha\mathbf{c}+\beta\mathbf{d})=\alpha\mathrm{L}(\mathbf{c})+\beta\mathrm{L}(\mathbf{d}). \quad \ldots\ldots\ldots\ldots\ldots \quad (4)$$

Substituting (2) and (3) in (1) we find, correct to linear terms in the electric field, that the components of **c** satisfy the equations

$$\mathrm{L}(c_1)+\frac{\partial c_1}{\partial t}-\Omega c_2=-\frac{\varepsilon\hbar}{m}F_x, \quad \ldots\ldots\ldots\ldots\ldots \quad (5)$$

$$\mathrm{L}(c_2)+\frac{\partial c_2}{\partial t}+\Omega c_1=-\frac{\varepsilon\hbar}{m}F_y. \quad \ldots\ldots\ldots\ldots\ldots \quad (6)$$

Now, by introducing a second imaginary j ($\mathrm{j}^2=-1$, $\mathrm{ij}\neq-1$) equations (5) and (6) may be combined into the single equation

$$\mathrm{L}(\mathsf{C})+\frac{\partial \mathsf{C}}{\partial t}+\mathrm{j}\Omega\mathsf{C}=-\frac{\varepsilon\hbar}{m}\mathsf{F}, \quad \ldots\ldots\ldots\ldots\ldots \quad (7)$$

where $\mathsf{C}=c_1+\mathrm{j}c_2$, and $\mathsf{F}=F_x+\mathrm{j}F_y$.

It is now convenient to introduce a formal relaxation time $\tau(\omega,\Omega)$ defined by

$$\mathsf{C}=-\frac{\varepsilon\hbar}{m}\mathsf{F}\tau(\omega,\Omega). \quad \ldots\ldots\ldots\ldots\ldots \quad (8)$$

Since $\mathsf{F}\propto e^{\mathrm{i}\omega t}$, equation (8) is a solution of (7) provided $\tau(\omega,\Omega)$ satisfies

$$\mathrm{L}[\tau(\omega,\Omega)]+(\mathrm{i}\omega+\mathrm{j}\Omega)\tau(\omega,\Omega)=1. \quad \ldots\ldots\ldots\ldots \quad (9)$$

With $\omega=\Omega=0$, equation (9) reduces to

$$\mathrm{L}[\tau(0,0)]=1, \quad \ldots\ldots\ldots\ldots\ldots \quad (10)$$

where $\tau(0,0)$ corresponds to one of the formal relaxation times introduced by Dingle (1956*b*).

Solution of the Boltzmann equation (9) is extremely difficult. The object of the present note is to direct attention to one interesting simplifying feature—that the solutions of (9) for the three field combinations ($\omega=0$, $\Omega\neq0$), ($\omega\neq0$, $\Omega=0$), and ($\omega\neq0$, $\Omega\neq0$) are essentially the same, so that solution of the Boltzmann equation for any one of the above cases in fact implies a full solution of the problem for all the cases.

Finally, it is of interest to compare the " universal " and the " formal " relaxation time theories. As mentioned above, the influence of a uniform magnetic field and a high frequency electric field on the transport properties may be obtained by replacing τ^{-1} by $\tau^{-1}+\mathrm{i}\omega+\mathrm{j}\Omega$. Analogously, for the formal relaxation time there is the replacement of the collision operator L by $\mathrm{L}+\mathrm{i}\omega+\mathrm{j}\Omega$.

The author is indebted to the Commonwealth Scientific and Industrial Research Organization for the award of a Studentship.

References

DINGLE, R. B. (1956*a*).—*Physica* **22**: 701.
DINGLE, R. B. (1956*b*).—*Physica* **22**: 698.
HOWARTH, D. J., and SONDHEIMER, E. H. (1953).—*Proc. Roy. Soc.* A **219**: 53.
WILSON, A. H. (1953).—"The Theory of Metals." 2nd Ed. (Cambridge Univ. Press.)

AUSTRALASIAN MEDICAL PUBLISHING CO. LTD.
SEAMER AND ARUNDEL STS., GLEBE, SYDNEY

NOTICE TO CONTRIBUTORS

1. General.—Papers will be considered for publication irrespective of the organization to which the authors are attached. Intending contributors, in preparing papers for submission, should follow the general style adopted in this Journal. A "Guide to Authors" may be obtained on application to the Editor.

   Papers, which should be written as concisely as possible, should be double-space typed on one side of the paper only and with liberal side margins. Every page, including those carrying tables and illustrations, should be numbered. Lengthy manuscripts should be accompanied by a table of contents with headings appropriately numbered. The organization to which the author is attached should be indicated in a footnote.

   The original typescript and one carbon copy should be forwarded.

2. Summary.—A short summary, which may be paragraphed, should precede the introduction to the paper. It should be of an informative character and written in such a way that it can be used by abstracting journals without amendment.

3. References.—References are to be cited in the text by the year of publication, e.g. Thomson (1948), and are to be arranged alphabetically, giving the author's name and initials followed by the year of publication, title of the paper (if desired), title of the periodical, volume, and pages, thus:

   Thomson, G. P. (1948).—*Proc. Phys. Soc. Lond.* **61**: 403–16.

   Abbreviations of titles of periodicals should conform to those used in "A World List of Scientific Periodicals".

4. Illustrations.—Line drawings should be made with indian ink on white drawing paper (preferably bristol board), tracing paper, or graph paper. Only blue or very pale grey ruled graph paper should be used as yellow, green, and red lines are difficult to screen out. Blue linen should not be used. Lines should be sufficiently thick and symbols sufficiently large to allow of reduction to a page size of 7¾ by 5 in. (or less). It is suggested that figures be drawn approximately twice the width at which they are to be printed. Lettering and numbering should only be pencilled on the drawings as the appropriate type will be substituted before the blocks are made. Original drawings should be submitted but, where possible, photographic copies or blue-prints of the originals should also be forwarded.

   Half-tone photographs, which should be included only where essential, should be on glossy paper and show as much contrast as possible.

5. Reprints.—Individual authors receive 40 copies of each paper free, two authors receive 30 each, and three or more authors receive 20 each, but additional copies may be purchased.

CONTENTS

PAGE

General Method of Exact Solution of the Concentration-dependent Diffusion Equation. By J. R. Philip 1

The Function inverfc θ. By J. R. Philip 13

The Distribution of Ions formed by Attachment of Electrons moving in a Steady State of Motion through a Gas. By C. A. Hurst and L. G. H. Huxley 21

The Design of Photographic Objectives of the Triplet Family. II. The Initial Design of Compound Triplet Systems. By F. C. Cruickshank 27

The Damping of Water Waves by Surface Films. By R. G. Vines .. 43

Sunrise and Eclipse Effects on the Ionosphere at Brisbane. By G. G. Bowman 52

A Relationship between Spread-F and the Height of the F_2 Ionospheric Layer. By G. G. Bowman 69

Thermal Transients associated with Natural Convection. By R. C. L. Bosworth and C. M. Groden 73

Attempts to measure the Inductive Element associated with the Natural Convection of Heat. By R. C. L. Bosworth 84

Short Communications

An Analogy between the Effects of Fluctuating Electric Fields and Steady Magnetic Fields in Isotropic Conductors when a Universal Relaxation Time cannot be Defined. By E. J. Moore 95

AUSTRALASIAN MEDICAL PUBLISHING CO. LTD.
SEAMER AND ARUNDEL STS., GLEBE, SYDNEY